Weak Epistasis May Drive Adaptation in Recombining Bacteria

Brian J. Arnold,*,†,1 Michael U. Gutmann,‡ Yonatan H. Grad,§ Samuel K. Sheppard, Jukka Corander,††,‡‡ Marc Lipsitch,*,†,§ and William P. Hanage*,†**

*Center for Communicable Disease Dynamics, †Department of Epidemiology, and §Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, Massachusetts 02115, ‡School of Informatics, University of Edinburgh, EH8 9AB, United Kingdom, **Department of Biology and Biochemistry, University of Bath, BA2 7AY, United Kingdom, ††Department of Biostatistics, University of Oslo, Blindern, 0317, Norway, and ‡‡Helsinki Institute for Information Technology HIIT, Department of Mathematics and Statistics, University of Helsinki, 00014 Finland

ORCID ID: 0000-0001-5646-1314 (Y.H.G.)

ABSTRACT The impact of epistasis on the evolution of multi-locus traits depends on recombination. While sexually reproducing eukaryotes recombine so frequently that epistasis between polymorphisms is not considered to play a large role in short-term adaptation, many bacteria also recombine, some to the degree that their populations are described as "panmictic" or "freely recombining." However, whether this recombination is sufficient to limit the ability of selection to act on epistatic contributions to fitness is unknown. We quantify homologous recombination in five bacterial pathogens and use these parameter estimates in a multilocus model of bacterial evolution with additive and epistatic effects. We find that even for highly recombining species (*e.g.*, *Streptococcus pneumoniae* or *Helicobacter pylori*), selection on weak interactions between distant mutations is nearly as efficient as for an asexual species, likely because homologous recombination typically transfers only short segments. However, for strong epistasis, bacterial recombination accelerates selection, with the dynamics dependent on the amount of recombination and the number of loci. Epistasis may thus play an important role in both the short- and long-term adaptive evolution of bacteria, and, unlike in eukaryotes, is not limited to strong effect sizes, closely linked loci, or other conditions that limit the impact of recombination.

KEYWORDS multilocus selection; bacteria; homologous recombination; epistasis; approximate Bayesian computation

EPISTASIS for fitness traits may arise from any nonlinearity in the multi-locus genotype-to-fitness map (Whitlock *et al.* 1995; Martin *et al.* 2007). The role of epistasis in adaptive evolution has been debated since the origin of population genetics (Fisher 1918; Wright 1931; Bulmer 1980; Coyne *et al.* 2000; Goodnight and Wade 2000; Carter *et al.* 2005; Hill *et al.* 2008; Crow 2010; Hansen 2013; Mäki-Tanila and Hill 2014). Evolutionary dynamics with epistasis are complex compared to single locus models, as the ability of interactions at the gene level to contribute to selection responses and adaptation at the population level depends on the relative amounts of recombination and epistatic effect sizes (Kimura 1965; Neher and Shraiman 2009), allele frequencies (Hill *et al.* 2008), nonequilibrium population dynamics (Goodnight 1988; Cheverud and Routman 1996; Barton and Turelli 2004; Hallander and Waldmann 2007), and the timescale under consideration (Yukilevich *et al.* 2008; Paixão and Barton 2016). In eukaryotic models, recombination can dominate the microevolutionary process as a result of linkage equilibrium between loci, making epistatic combinations of alleles less heritable across generations unless epistatic effects are strong (Kimura 1965; Neher and Shraiman 2009). However, genetic drift and selection may change allele frequencies over long periods of evolutionary time, homogenizing genetic backgrounds and allowing effects that are epistatic at the gene level to contribute to the additive genetic variance between individuals, and, thus, the long-term response to selection (Hallander and Waldmann 2007; Paixão and Barton 2016).

While these models have shown that recombination affects selection on epistatic interactions, epistasis may also affect the

doi: https://doi.org/10.1534/genetics.117.300662
Manuscript received October 10, 2017; accepted for publication January 1, 2018; published Early Online January 12, 2018.
Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.117.300662/-/DC1.
[1]Corresponding author: Center for Communicable Disease Dynamics, Harvard T. H. Chan School of Public Health, 677 Huntington Ave., Suite 506, Boston, MA 02115. E-mail: barnold@hsph.harvard.edu

ability of recombination to drive adaptation over long periods of evolutionary time by changing patterns of linkage via selection. Negative and positive epistasis produce linkage disequilibrium (LD) of the same sign (Eshel and Feldman 1970), but only positive LD generates additive variance in fitness and enhances the efficacy of natural selection (Fisher 1930; Bulmer 1976). Since recombination destroys correlations (positive or negative) between mutations, it only increases LD and variance in fitness when linkage is negative (Charlesworth 1993; Barton 1995; but see Barton 2010 for a review).

These studies have given us enormous clarity on the complex effects epistasis may have on evolution, but they have focused on eukaryotes and are less applicable to organisms that do not reproduce sexually. Bacteria, which have colonized almost every conceivable ecological niche, also recombine to varying degrees through multiple mechanisms, leading to a continuum of genealogical structures from clonal to "fully sexual" (Smith *et al.* 1993). For microbes that recombine in moderation (*e.g.*, *Staphylococcus aureus*), it is clear that selection can easily act on epistatic allele combinations: mutations will likely exist only in the genetic background on which they arose, such that any background-specific epistatic effects are heritable through time and may spread through populations via selection. However, some bacteria recombine at much higher rates, to an extent that they have been historically labeled as "fully sexual" or "freely recombining" (Smith *et al.* 1993; Suerbaum *et al.* 1998). We do not know whether such highly recombining species decouple interacting mutations frequently enough to prevent selection on *weak* epistatic effects, or how strong epistasis must be to dominate the microevolutionary process and drive adaptation in these species.

The most widely studied models of selection response and adaptation, those from quantitative genetics of eukaryotes, have not been directly applicable to bacteria (or Archaea; Levin and Bergstrom 2000), especially since even partial linkage between loci prevents the partitioning of genetic variance in a population into additive and epistatic components (Hill and Mäki-Tanila 2015). While a few studies have explored the effects of bacterial recombination on adaptation (Cohen *et al.* 2005; Cooper 2007; Levin and Cornejo 2009), even fewer explicitly incorporate epistasis (but see Moradigaravand and Engelstädter 2012). Consequently, we know little about how epistasis affects adaptive processes and the evolution of multilocus phenotypes such as antibiotic resistance, antigenic profile, or metabolic output, all of which likely involve epistatic interactions (Gupta *et al.* 1996; Trindade *et al.* 2009; He *et al.* 2010). Answering these questions, which are relevant for the entire bacterial and Archaeal kingdoms of life, requires multilocus models with epistasis that account for the features of bacterial recombination that involves the transfer of smaller DNA segments, together with accurate estimates of genome-wide recombination parameters.

We use a multi-locus model of bacterial evolution to study how neutral patterns of LD affect selection on standing genetic variation when both additive and epistatic effects contribute to fitness differences between individuals. In this model, mutations are selected from neutral mutation-recombination-drift balance. We then vary the magnitude and genetic basis of fitness differences but use biologically realistic levels of bacterial recombination by inferring these parameters from genomic data of five pathogens (*Staphylococcus aureus*, *Campylobacter jejuni*, *Streptococcus pneumoniae*, *Neisseria gonorrheae*, and *Helicobacter pylori*), using Approximate Bayesian Computation (ABC) and machine learning. The bacteria we chose exhibit strikingly different degrees of genome-wide linkage and include some of the most highly recombining bacteria known. Despite large variation in recombination rates among species, we find that selection responses in bacteria are nearly as strong as in asexuals in the presence of weak epistatic interactions ($N|s_i| \approx 3$–10, where s_i is the epistatic effect per locus-pair) regardless of their physical proximity on a circular chromosome, even for highly recombining pathogens that have been previously labeled as freely recombining (Suerbaum *et al.* 1998). As the strength of epistasis $N|s_i|$ increases, recombining bacteria transition into a regime where they adapt faster than asexuals. However, less-recombining bacteria respond more strongly to selection on simple traits (three loci) whereas highly recombining bacteria have stronger responses for more complex traits (10 loci). Given the wide range of recombination rates observed in nature, this may have broad implications for the ways in which bacteria respond to different selective pressures.

Materials and Methods

Genomic data

We analyzed previously published genomic datasets (Supplemental Material, Table S1 in File S1). Using an amino acid file from a reference genome (Table S2 in File S1), we annotated *de novo* assemblies from each species with PROKKA (Seemann 2014), and identified core and accessory genes with ROARY (Page *et al.* 2015). Only core genes—present in all samples in a species—that were also present in the reference genome were used for downstream analyses. All position information between genes and polymorphic sites is derived from their relative positions in the reference genome used to annotate *de novo* assemblies, not from a reference-based DNA alignment (Figure S2 in File S1). For analyses that required polarized mutations, we used Mauve (Darling *et al.* 2004) to align these reference genomes to an outgroup species (Table S4 in File S1) to infer the derived and ancestral state of each polymorphism.

Subsample selection

For datasets with a wide geographic or temporal distribution, we partitioned samples by geography and collection date into smaller subsamples to minimize the effects of sampling and structure on population genetic parameter estimates. Subsamples consisted of isolates from a similar geographic region (to avoid genetic isolation by distance) and also had similar collection dates, since serial samples can skew genealogies to

have longer terminal branches, potentially leading to substantial overestimates of LD or related statistics (Slatkin 1994). Subsamples that had the least evidence of substructure (near-zero estimates of Tajima's D) were chosen for analysis (Table S2 in File S1).

Estimation of population genetic parameters

Summary statistics: We used five summary statistics to fit coalescent models to observed genomic data: the minimum number of mutations per site (Tajima 1996) to estimate the population mutation rate $\theta = 2N\mu$, and four recombination-related statistics to estimate both the population recombination rate $\rho = 2Nr$ and the mean of the geometrically distributed DNA tract lengths transferred between donor and recipient ($1/q$, where q is the geometric distribution parameter). We used pairwise compatibility (PC) to quantify the amount of recombination that has taken place between two SNPs, which are compatible with an infinite-sites model of no recombination if <4 haplotypes are observed; either recurrent mutation, or, more likely, recombination gives rise to four observed haplotypes (Figure S1 in File S1). PC generally decreases as a function of the genomic distance between SNPs in recombining bacteria, and the shape of this decay contains information about both ρ and q. Consequently, for four different inter-SNP distance categories, we calculated mean PC to capture both short- and long-range recombination dynamics (similar to Ansari and Didelot 2014). We calculated mean PC only for synonymous, intermediate-frequency SNPs (10–90% frequency in sample), and the minimum number of mutations only for fourfold degenerate sites.

Since we are interested in inferring parameters genome-wide, we compared observed summaries from k genomic windows with k simulated datasets, since our coalescent model could only simulate segments of maximal length roughly equal to 150 kb (below). To find the right set of summary statistics for inference of ρ and q, we simulated a k-window dataset with known parameter values, and compared summaries calculated from this "true" dataset with those calculated from k-window datasets simulated across a grid of θ, ρ, and q values. Specifically, for each simulated dataset in the grid, we calculated a discrepancy between k simulated and "true" summary statistic values using a Kolmogorov-Smirnov statistic, one for each of five summaries (above). We summed these five Kolmogorov-Smirnov statistics for each point in the grid (Figure S3 in File S1). When we measured PC at two short-range inter-SNP distances (with respect to the decay of PC *vs.* distance) and two long-range inter-SNP distances, datasets simulated with parameter values close to "true" values had the smallest discrepancy. Thus, we chose different inter-SNP distances and window lengths for each species (Table S6 in File S1) based on the decay of PC *vs.* distance and on the computational resources needed to simulate data within the parameter space, as high ρ required more memory or time, depending on the simulator used. We found that comparing the mean values only of PC summaries between simulated and observed data, as opposed to the full distribution of k values, resulted in a reduced ability to distinguish between datasets simulated with high ρ, large q (small tract lengths) and low ρ, small q (large tract lengths).

Coalescent simulations: Using CoaSim (Mailund *et al.* 2005), we constructed a novel, finite-sites coalescent model to simulate genomic DNA, which was required to accurately model sequence alignments from highly diverse pathogens like *H. pylori* that frequently have multiple bases per site. A finite-sites model not only enables more accurate inference of $\theta = 2N\mu$ but also more precise estimates of recombination parameters since back mutation mimics recombination by affecting PC, particularly for species with high transition: transversion ratios (Ti:Tv). Our coalescent model thus accounted for species-specific base compositions and Ti:Tv, which we estimated from a reference genome or fourfold degenerate sites, respectively (Table S6 in File S1). For less diverse species, we used ms (Hudson 2002) to simulate longer DNA sequences, which yielded better estimates of recombination parameters as there were more SNP pairs with particular inter-SNP distances. While previous studies have only used a single coalescent simulator (De Maio and Wilson 2017), we used two simulators to exploit the benefits of each: ms is computationally efficient, but Coasim allows specification of more complicated mutation models.

Parameter estimation: We constructed a novel approach to estimate population genetic parameters in bacteria, based on the statistical framework of Gutmann and Corander (2016). Our method uses Bayesian Optimization with Gaussian Process regression (GPR) to model the discrepancy between simulated and observed datasets. We chose this machine-learning approach to prudently explore parameter space due to the computational requirements of the finite-sites coalescent simulator. For each set of parameters (θ, ρ, q) chosen by the method, we simulated k genomic windows, calculated five summary statistics (above) for each window, and used a Kolmogorov-Smirnov statistic to calculate five discrepancies between each set of k simulated and observed summaries. The final discrepancy was a sum of these five Kolmogorov-Smirnov statistics. We ran the inference algorithm for 320 iterations, simulating k windows each time (Figure S12 in File S1). The results stabilized after 200 iterations when the GPR no longer changed with additional acquisitions (visual inspection). The GPR model of the discrepancy was then used in an approximate Bayesian computation (ABC) framework to approximate the intractable likelihood function that enabled us to compute the posterior distributions by standard sequential Monte Carlo sampling (Gutmann and Corander 2016). Specifically, the likelihood function was approximated by the probability to draw discrepancy values from the GPR model that were less than a small threshold. Following common practice, the threshold was set to the 1% quantile of the discrepancy values of the simulated data. When given "observed" data that we generated with known parameter

values, our approach accurately estimated the input parameter values (Figure S4 in File S1).

Comparing k simulated DNA segments with k observed genomic windows implicitly assumes independence among windows, since each coalescent simulation is independent. To test whether this assumption affects parameter estimation, we simulated a large 200 kb DNA segment with a modest recombination rate and broke this segment up into 10 windows of 20 kb. We then simulated a grid of θ, ρ, and q values around the "true" values used to simulate the large segment, and calculated the discrepancy between simulated and "true" summaries, as above. Like before, datasets simulated with parameter values close to "true" values had the smallest discrepancy (Figure S3 in File S1), showing that this independence assumption should not greatly affect our parameter estimates and conclusions, at least within the range of recombination rates studied here.

Table 1 ABC parameter estimates

Species	Parameter	MDE	Mean (Lower CI/Upper CI)
S. aureus	θ	31.9	32.0 (24.6/41.6)
(n = 30)	ρ	11.5	12.2 (6.9/22.2)
	$1/q$	68	63 (30/120)
C. jejuni	θ	10.6	10.6 (7.6/14.4)
(n = 23)	ρ	0.6	0.6 (0.3/1.0)
	$1/q$	1429	1429 (667/3333)
S. pneumonia	θ	26.9	26.7 (19.8/35.5)
(n = 280)	ρ	11.5	11.3 (9.2/13.9)
	$1/q$	588	625 (455/833)
N. gonorrheae	θ	4.3	4.9 (1.4/19.5)
(n = 19)	ρ	8.6	8.2 (4.5/14.5)
	$1/q$	2500	3333 (1250/10000)
H. pylori	θ	133.7	129.0 (106.1/160.5)
(n = 21)	ρ	471.7	484.7 (378.8/652.7)
	$1/q$	49	49 (27/96)

MDE represents the minimum discrepancy estimate from the GPR, or the parameter value set with the smallest discrepancy between simulated and observed data. Estimates of θ and ρ are in units per kilobase, while $1/q$ is in base pair.

Epistasis simulations

Sign epistasis model: We created two Wright-Fisher forward-time simulators in C++ to simulate bacterial populations under neutrality and use these as starting conditions for simulations with selection. We first simulated five bacterial populations under neutral mutation-recombination-drift balance with a population size of $N = 1000$ for $10N$ generations, using population recombination and mutation rates that satisfied the parameter estimates for the five species in Table 1 (*e.g.*, we used recombination rate $r = 5.75 \times 10^{-6}$ per generation per base pair to simulate dynamics for *S. pneumoniae* so that $2Nr = 11.5$ per kilobase). While most bacteria likely have effective population sizes larger than $N = 1000$, evolutionary dynamics are largely controlled by the product of N and per-generation parameters such as mutation (μ) and recombination (r) rates and selective effects (s) (see "*Rescaling simulations*" below). We used a custom infinite-sites simulator for most of these neutral simulations in which bacterial recombination was modeled as gene conversion, with ~Poisson($\rho X/2N$) recombination events per individual per generation, where ρ is the population recombination rate per site, and X is the total number of simulated sites in base pair. Each recombination event transferred a geometrically distributed DNA tract from a randomly selected chromosome, which served as the donor for all recombination in the case of multiple events. These recombination events only changed the multi-locus genotype of the recipient chromosome if the coordinates of the donor DNA tract overlapped with positions of any polymorphic loci in the recipient, and the donor and recipient had different alleles at these loci. For computational efficiency, the number of simulated sites (X) differed for each species such that we could randomly select at least 10 polymorphic sites separated by distances of at least two times the mean recombination tract length (Table 1). We thus simulated longer fragments for species that exchange longer DNA tracts. However, we used SFS_CODE (Hernandez 2008) to simulate DNA for *H. pylori*, since SFS_CODE has a finite-sites model appropriate for species with very high mutation rates, and to simulate crossover recombination for eukaryotic DNA (to model budding yeast).

From these neutral simulations, we randomly selected L polymorphic loci, represented as $l_k = \pm 1$ where $k = 1, \ldots L$. We selected these loci to be separated by genetic distances of at least two times the estimated mean recombination tract length (Table 1), since linkage does not decay further for loci separated by distances longer than the mean tract length. These loci thus exhibit "loose" linkage dynamics representative of those that are distantly spaced, and they follow a U-shaped frequency distribution if the derived allele is randomly assigned at each locus, with -1 and $+1$ being the ancestral and derived allele, respectively. We devised this method to initiate simulations of selection with L loci because starting levels of LD and allele frequencies affect the additive and epistatic genetic variances of the population, and thus the potential responses to selection (Mäki-Tanila and Hill 2014; Hill and Mäki-Tanila 2015). For each parameter set listed in Table 1, we simulated 400 replicates, and, for each replicate, we randomly sampled L loci five times since we simulated large DNA segments that accumulated many polymorphic loci at different frequencies by $10N$ generations. Thus, we had 2000 starting conditions per parameter set.

To model multilocus selection, we assign each locus an additive effect (s_a) and each locus pair an epistatic interaction effect (s_i). For a given simulation, s_a and s_i are the same for each locus or locus pair, respectively. The fitness of each individual is calculated similar to Neher and Shraiman (2009) as

$$\text{Fitness} = 1 + s_a \sum_{k}^{L} l_k + \sum_{k<j} s_i l_k l_j I_{jk},$$

where we introduce I as a random variable that may be 1 or -1 with equal chance, which changes the sign of s_i such that there was no tendency for epistasis to be positive or negative. Distributions of epistasis with zero mean have been observed in bacteria and other microbes (Martin *et al.* 2007).

Consequently, stronger epistatic effects cause the fitness landscape to have multiple, steeper peaks (see Weinreich *et al.* 2005; Figure 2B) for an example with two loci) We modeled a generation of bacterial evolution by sampling chromosomes with replacement according to their relative finesses, such that a randomly sampled chromosome with fitness w_i was passed to the next generation with probability w_i/w_{max}, where w_{max} is the fitness of the most fit genotype in that generation. We repeated this process until N chromosomes were retained, and each chromosome was allowed to recombine with a single, randomly selected donor individual. As with the neutral simulations (above), the number of recombination events per individual was $\sim$Poisson($\rho X/2N$), and each event transferred a geometrically distributed DNA tract. Recombination events changed only the multi-locus genotype of the recipient chromosome if the coordinates of the donor DNA tract overlapped with positions of any polymorphic loci under selection (here $L = 3$ or 10) and the donor and recipient had different alleles at these loci. As noted above, the positions of these L loci came from the neutral simulations. Likewise, eukaryotic evolution was modeled by sampling N linear chromosomes with replacement according to relative fitness, each time recombining this chromosome with another, randomly selected chromosome according to the number of crossover breakpoints. Here, we model multi-locus selection in terms of per locus additive effects and per locus pair epistatic effects, as opposed to the population level additive and epistatic genetic variances that are classically used in quantitative genetics, because these population level quantities can only be reliably calculated if loci are in linkage equilibrium (Hill and Mäki-Tanila 2015), a condition that is not met by most bacteria. Source code for this simulator is freely available (github.com/brian-arnold/BacteriaEpistasisSimulator).

Positive and negative epistasis model: We used a slightly modified simulation framework to model strictly positive or negative epistasis that was similar to the above framework but with two key differences: (1) the L polymorphic loci were represented as $l_k = 0$ or 1 and (2) I was no longer a random variable but a constant, either 1 for positive epistasis or -1 for negative epistasis. Due to these modifications, epistasis only occurred between beneficial alleles and not between any pair of loci in which at least one allele was null (the 0 allele).

Analysis of simulations: We ran each simulation until a certain number of generations passed, using $0.2N$ (or 200) generations to assess short-term evolution and $10N$ (or 10,000) generations to assess long-term evolution in which all polymorphic sites fixed for one allele. At these stopping points, we calculated the standardized response to selection

$$R = \frac{\overline{W_{AS}} - \overline{W_{BS}}}{\sigma_{BS}},$$

or the difference between the mean population fitness before ($\overline{W_{BS}}$) and after ($\overline{W_{AS}}$) selection, where σ_{BS} is the variance in fitness before selection. We also ran an asexual "control" with no recombination, and calculated the same response to selection (R_{asex}). We used these two quantities to calculate the relative speed of adaptation R/R_{asex} and study how clonal these pathogens behave in the presence of varying amounts of epistasis.

Rescaling simulations: We inferred both the population mutation rate $\theta = 2N\mu$ and population recombination rate $\rho = 2Nr$ from genomic data, but an exact population size must be specified for our forward-time simulations with epistasis. The rescaling of forward-time simulations has been previously explored for simple scenarios of selection (Hoggart *et al.* 2007), showing that only the products of $N\mu$, Nr, and Ns matter, such that one may choose smaller N for computational efficiency and increase μ, r, and s accordingly. Thus, one may simulate a rescaled population with N/λ, $\lambda\mu$, λr, and λs, where $\lambda > 1$ represents some rescaling factor (Hoggart *et al.* 2007). We confirm that rescaling preserves the population genetic dynamics of more complex selection with sign epistasis and recombination as long as the ratio of r/s is conserved. For instance, simulating a population size of $N = 10{,}000$ or $N/\lambda = 1000$ ($\lambda = 10$) gives remarkably similar results on different evolutionary timescales (different by a factor of λ), as long as r and s are changed accordingly such that Nr, Ns, and r/s are constant (Figure S13 in File S1).

Data availability

All data used in this study have been previously published and are listed in Table S1 in File S1: *Staphylococcus aureus* (SRA PRJEB2478), *Campylobacter jejuni* (dryad doi: 10.5061/dryad.28n35), *Streptococcus pneumoniae* (ENA ERP000809), *Neisseria gonorrheae* (ENA study accession PRJEB7904), and *Helicobacter pylori* (NCBI see publication in Table S1 in File S1 for accession numbers).

Results

Recombination parameter estimates

In order to simulate bacterial evolution with selection and biologically realistic recombination rates, we first inferred recombination parameters using five genomic datasets from *S. aureus*, *C. jejuni*, *S. pneumoniae*, *N. gonorrheae*, and *H. pylori*. We inferred both the rate of DNA transfer and the mean tract lengths involved using a approach that we developed (below), as opposed to using previously published estimates, because other popular recombination-detection programs have known biases toward detecting larger recombination events between more diverged sequences (Croucher *et al.* 2014). While these programs may miss short recombination events or transfers between less diverged sequences, these events affect PC and LD such that use of these summary statistics facilitates parameter inference in species that have less diversity (*e.g.*, *N. gonorrheae*) or exchange short DNA tracts. Consequently, methods that use correlations between mutations to quantify recombination have

been gaining popularity (Ansari and Didelot 2014; Lin and Kussell 2016; De Maio and Wilson 2017). While the approach of Lin and Kussell (2016) is unique, our method is similar to that of others (Ansari and Didelot 2014; De Maio and Wilson 2017) but differs in the summary statistics used to compare simulated and observed datasets in order to accurately infer both the rate and mean lengths of DNA transfers, which have critical implications for how selection acts on epistatic interactions.

For genomic datasets containing isolates collected across many years or large geographic areas, we selected a restricted subsample (Table S1 in File S1) to avoid population structure that could confound estimates of recombination. Analysis of samples taken at very different time points may artificially elongate terminal branches of the genealogy, leading to underestimates of LD (and related statistics) and overestimates of recombination (Slatkin 1994). On average, each dataset had over 1000 core genes containing almost 1 Mb of DNA (Table S2 in File S1). We used fourfold degenerate sites in each sample to calculate Tajima's D, which was typically near zero (Table S2 in File S1), suggesting these samples came from populations that have not experienced nonequilibrium demography, and are not strongly structured, both of which may affect estimates of recombination parameters. More information on data processing can be found in the *Materials and Methods*.

Using Approximate Bayesian Computation (ABC) coupled with Bayesian Optimization (Gutmann and Corander 2016), we fit customized coalescent models with gene conversion to summaries of genomic data in order to infer three parameters: the population mutation rate $\theta = 2N\mu$, the population recombination rate $\rho = 2Nr$, and the mean of DNA tract lengths transferred between donor and recipient [$\sim$Geometric(q), where $1/q$ is the mean tract length]. To summarize the statistical associations between single-nucleotide polymorphisms (SNPs), we developed an approach that gave accurate estimates of ρ and q (*Materials and Methods*). Briefly, we used PC to quantify the amount of recombination that has taken place between two SNPs, which are compatible with a single phylogeny if less than four haplotypes are observed (Wilson 1965); either recurrent mutation, or, more likely, recombination, gives rise to four haplotypes (Figure S1 in File S1). PC is equivalent to the four-gamete test (Hudson and Kaplan 1985) and quantifies historical recombination similar to measures of LD, such as D' or r^2 (Slatkin 2008). We quantified PC within k genomic windows, and compared these observed estimates to k simulated windows via a Kolmogorov-Smirnov statistic, which captures higher moments of the PC distribution (see Figure S2 in File S1 for a diagram of our analysis). To infer θ, we calculated the minimum number of mutations per site (Tajima 1996), a sufficient statistic for this parameter (Roychoudhury and Wakeley 2010). With these summary statistics, input recombination parameters were accurately recovered from simulated datasets (Figures S3 and S4 in File S1).

Recombination estimates for the five bacteria studied are summarized in Table 1, and simulations with these parameters largely fit observed data (Figure 1). While our parameter estimates for some species (*S. pneumoniae* and *C. jejuni*) are generally consistent with previous work, we observe some differences for other species that could be relevant to selection on epistatic interactions. For instance, parameter estimates for *H. pylori* revealed an extremely high value of $\rho = 472$ per kilobase but short tract lengths $\sim$50 bp that are approximately an order of magnitude smaller than previous estimates of $\sim$400 bp from genomic data (Falush *et al.* 2001; Kennemann *et al.* 2011), yet in agreement with short lengths reported from *in vitro* experiments that used diverged donor and recipient strains (Bubendorfer *et al.* 2016). We also find that *S. aureus* frequently transfers ($\rho = 11.5$ per kb) tracts $\sim$70 bp in length, which are also an order of magnitude smaller than previous estimates of $\sim$650 bp (Méric *et al.* 2015). Nonetheless, *S. aureus* still exhibits high genome-wide linkage, since these small transfers affect few SNPs. For both species, these short tract lengths agree with PC decaying within 100 bp (Figure 1), as the PC *vs.* distance distribution is expected to asymptote near the mean tract length because SNPs separated by larger distances are equally likely to be unlinked. Such short recombination events have important consequences for evolution (below).

To our knowledge, this analysis is the first to infer both a recombination rate ($\rho = 8.6$ per kilobase) and mean tract length in *N. gonorrheae*, which appear to transfer long segments ($\sim$2.5 kb) since PC decays slowly with distance (Figure 1). The PC data are notably noisy (in particular the change that occurs among SNP pairs separated by $\sim$500–700 bp), which may be due to rearrangements that have occurred since the divergence of our sample and the reference sequence used to estimate inter-SNP distances. Results are similar when we use a different reference sequence (Figure S5A in File S1), suggesting the rearrangement may be a derived feature of our sample. While we do not know the exact effect this noise has on our parameter estimates, it may have led to a slight overestimate of recombination rates, as mean PC from simulations parameterized with minimum discrepancy estimate (MDE) values listed in Table 1 is lower ($\sim$10%) than mean PC between randomly sampled SNP pairs, irrespective of distance (Figure S5B in File S1).

We thus have within our dataset parameter estimates from a diverse set of bacteria that represent the many ways bacteria may transfer DNA, including very high or low rates of transfer with short or long tract lengths. A full description of the parameter inference results can be found in the Supplemental Text in File S1.

Multi-locus simulations of adaptation

With these recombination parameters, we used simulations of bacterial evolution to study how epistasis may affect the short-term rate of adaptation, modeling L polymorphic loci, distantly spaced on a circular chromosome. These are drawn from mutation-recombination-drift balance, follow a neutral U-shaped frequency distribution (the beneficial allele is randomly assigned at each locus when selection starts to occur),

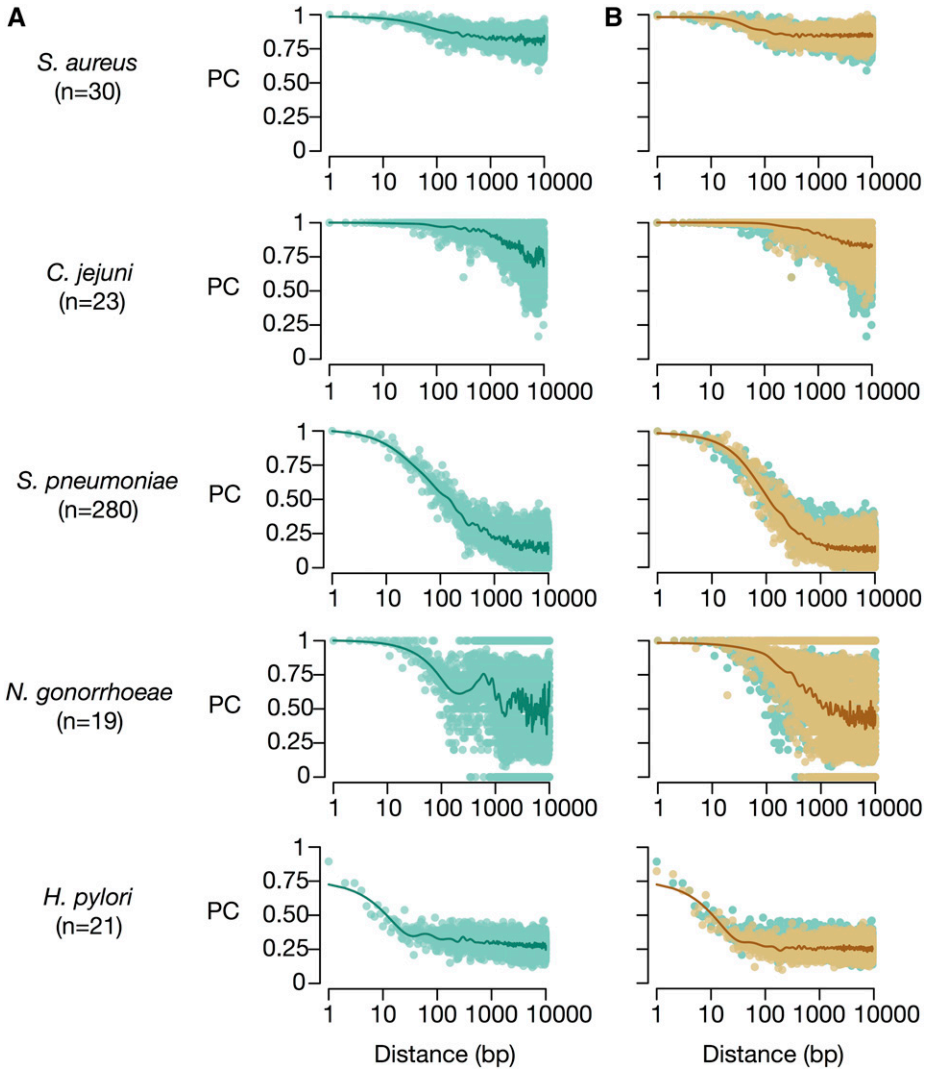

Figure 1 Observed pairwise compatibility *vs.* distance. (A) Patterns of PC (green) vary among the species included in this study. Since PC measures the compatibility of two SNPs with a single phylogeny, these data indicate that SNPs $>$ ~1kb apart have distinct phylogenetic histories from recombination, with the exception of *S. aureus* which exhibits linkage. (B) Simulated patterns of PC *vs.* distance (brown) using parameter estimates from Table 1 fit observed data well. We note that the sensitivity of PC to sample size (n) makes these patterns not directly comparable across species, and that the product of effective population size and recombination (or $\rho = 2Nr$) affects PC.

and exhibit pairwise LD according to the levels of recombination in Table 1. We specifically chose these starting conditions because initial allele frequencies and LD affect selection responses, since extreme allele frequencies and LD "convert" epistatic genetic variance to additive genetic variance (Mäki-Tanila and Hill 2014; Hill and Mäki-Tanila 2015). For these L loci, we vary the additive fitness effects of each locus and the epistatic fitness effects of each locus pair. The model uses sign epistasis, with effect sizes randomly assigned as positive or negative such that the mean effect is zero (*Materials and Methods*). Populations were evolved for a short or long period of evolutionary time ($0.2N$ or $10N$ generations, respectively). At this point, we recorded patterns of LD and population fitness relative to an asexual control with the same fitness effect sizes to directly compare our results to those expected under clonal evolution. For contrast, we also modeled a linear eukaryotic chromosome with a relatively low crossover recombination rate equivalent to facultatively sexual yeast (Table S3 in File S1).

For a multi-locus trait with a complex fitness landscape, controlled by 10 loci, simulations with eukaryotic recombination rates had diminished short-term selection responses relative to an asexual, particularly for weak to intermediate pairwise epistatic effects ($N|s_i| = 0.1$–10; Figure 2A). These results were similar whether loci had additive effects that were weak or intermediate in strength ($Ns_a = 1$ or 10, where s_a is the additive effect), showing that even low levels of crossover recombination may antagonize adaptation. For reference, theory from single-locus population genetics has shown selection efficiently acts on mutations only when $N|s| >> 1$ (Ohta 1976). In contrast with the eukaryote, all the bacterial simulations examined had similar or slightly greater selection responses compared to an asexual when epistatic effect sizes were weak ($N|s_i| = 0.1$–3; Figure 2A), and these responses became much greater for intermediate to strong epistatic interactions ($N|s_i| = 6$–60). With strong epistatic effects, bacterial simulations with more recombination had faster rates of adaptation (Figure 2A), while adaptation rates for bacteria with less recombination were closer to the asexual rate. This trend was also apparent when loci had stronger additive effects. A hallmark of selection on epistatic interactions is increased LD measured as D' between

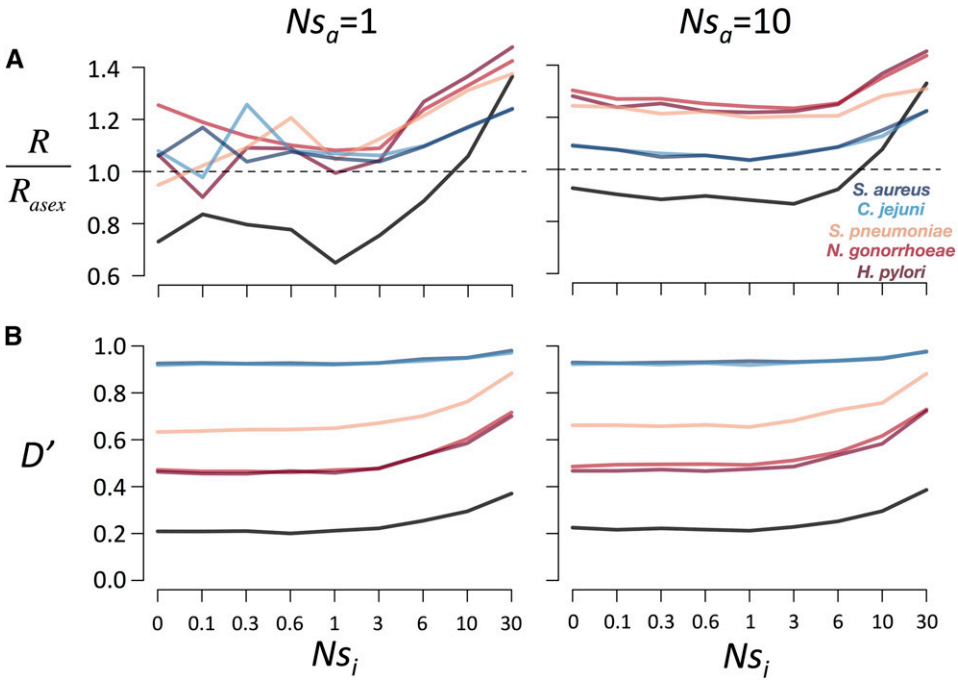

Figure 2 Higher recombining bacteria may have greater selection responses when multilocus epistatic traits are controlled by many loci (L = 10). (A) Selection responses of simulations parameterized by bacterial recombination rates, relative to an asexual control (R/R_{asex}), for increasing pairwise epistatic effects (Ns_i) when additive effects per locus are weak (Ns_a = 1; left) or strong (Ns_a = 10; right). (B) LD, as measured by the mean D′ across all locus pairs. For each plot, the mean of 2000 simulations is shown for each parameter set, and selection responses were calculated after 0.2 N generations of selection. Simulations were parameterized with values for *S. aureus* (dark blue), *C. jejuni* (light blue), *S. pneumoniae* (orange), *N. gonorrheae* (light red), *H. pylori* (dark red), and a eukaryote (black).

locus pairs, although weaker epistatic interactions ($N|s_i| \approx 3$) did not alter patterns of D' much from equilibrium levels despite shifting population fitness by ~1 SD (Figure S6 in File S1).

For a simpler multi-locus trait (and thus simpler fitness landscape) controlled by only three loci, we used larger selective effects to model the same total amount of selection on the trait, ensuring that $L \times Ns_a$ and $L \times (L-1) \times Ns_i/2$ were the same for both traits. For instance, per locus additive effects of $Ns_a = 10$ across 10 loci may similarly be modeled as $Ns_a = 33.3$ across three loci. Eukaryotic simulations exhibited diminished short-term selection responses when epistatic effects were weak to intermediate in strength ($N|s_i| = 1$–10), while bacterial simulations had similar or slightly greater responses to those of the asexual (Figure 3A). These dynamics are similar to the 10 locus results with weak epistasis ($N|s_i| < 3$), along with the trend that bacterial simulations began to adapt much faster than the asexual when epistatic effects were strong ($N|s_i| > 10$; Figure 3A). However, the selection dynamics became qualitatively different between a simple and complex trait under strong epistatic selection, as simulations of three loci with higher bacterial recombination had lower selection responses, perhaps because more fit allelic combinations were disrupted by recombination and unable to spread through populations. The maximum observed fitness of an individual within a population was generally higher for simulations with high recombination (Figure S7 in File S1), and we interpret our findings as showing that this effect (greater exploration of genotypic space) dominates for the complex trait, such that more recombination leads to greater adaptation (Figure 2), while the counteracting effect of recombination breaking favorable combinations (generating "recombination load") dominates in the simpler trait, such that lower (but nonzero) recombination rates maximize adaptation (Figure 3). For the simpler trait, with higher additive variance (Figure 3A, right) no pattern was apparent in the effect of recombination rates on rate of adaptation.

The ability of recombination to create more haplotype diversity before selection may explain part of the reason for these enhanced selection responses compared to an asexual, since selection started with loci under mutation-recombination-drift balance. Studying selection *only* on the standing genetic variation present at the onset of selection, by suppressing recombination during selection, shows that simulations with more recombination and haplotype diversity generally had greater selection responses (Figure 4). This trend was clearer for the 10-locus trait. However, starting conditions cannot account for the entirety of these enhanced selection responses, as rarely recombining bacteria for the three-locus trait, and highly recombining bacteria for the 10-locus trait, both had greater responses when recombination occurred during selection (Figure 2A and Figure 3A). For the case of the three-locus trait, highly recombining bacteria recombined frequently enough to antagonize adaptation, but selection responses were still greater than in asexuals.

The results shown here highlight the potential ability of bacteria to more rapidly respond to selective pressures than asexuals using weak epistatic interactions on the timescale of 0.2 N generations. Recombination may also shape long-term responses, particularly for weak epistatic effects ($N|s_i| = 1$–10) in which genetic variation persists in populations for longer time periods and recombination allows further exploration of the fitness landscape (Figure S8 in File S1). These delayed selection responses are particularly evident for the 10-locus

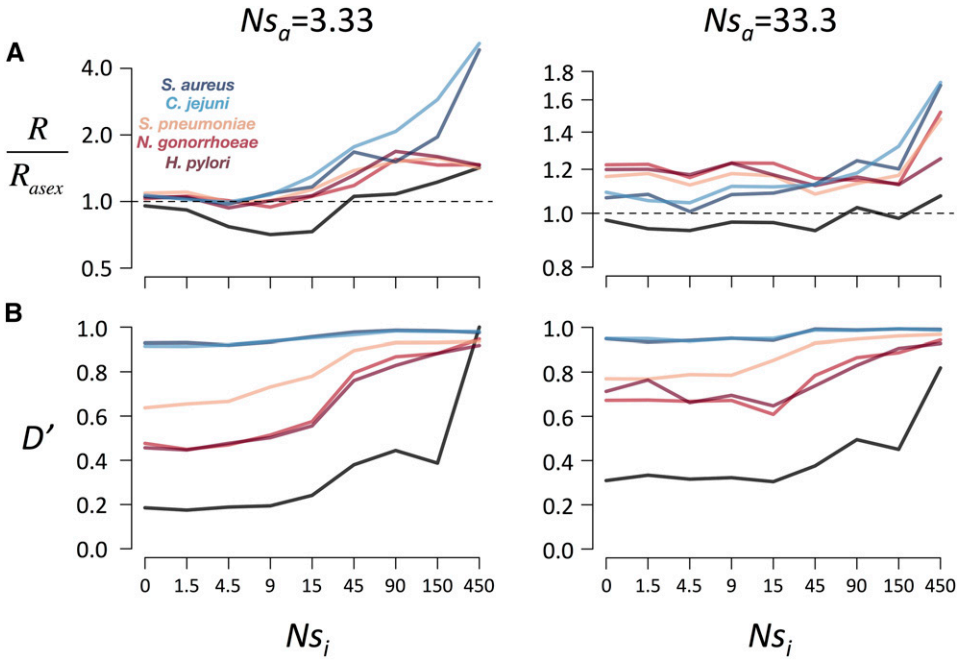

Figure 3 Bacterial recombination rates do not hinder short-term responses to selection for multilocus epistatic traits composed of few loci (L = 3). (A) Selection responses of simulations parameterized by bacterial recombination rates, relative to an asexual control (R/R_{asex}), for increasing pairwise epistatic selection effects (Ns_i) when additive effects per locus are weak (Ns_a = 1; left) or strong (Ns_a = 10; right). (B) LD, as measured by the mean D′ across all locus pairs. For each plot, the mean of 2000 simulations is shown for each parameter set, and selection responses were calculated after 0.2 N generations of selection. Colors follow the same scheme as Figure 2.

trait when additive effects are weak or intermediate in strength (Ns_a = 1 or 10; Figure S8 in File S1). For the more highly recombining simulations, especially the eukaryote, these delayed selection responses may be driven by epistatic genetic variance being "converted" to additive genetic variance as allele frequencies inevitably take on extreme values via drift and weak selection, preventing recombination from altering allele combinations (Paixão and Barton 2016). We would like to note that while eukaryotic recombination is predicted to enhance selection on beneficial alleles with no epistasis (Ns_a = 10, Ns_i = 0) by reducing clonal interference experienced in asexual populations (Barton and Otto 2005), we only observe this effect over longer periods of evolutionary time beyond 0.2 N generations (Figure S8 in File S1). This observation is likely influenced by our starting conditions of polymorphisms in mutation-recombination-drift balance that exhibit a U-shaped distribution, with neutral mutations at intermediate frequencies occurring on multiple genetic backgrounds. The benefit of recombination is particularly evident for more complex traits (L = 10) in which mutation and drift alone do not generate as much haplotype diversity.

In addition to modeling sign epistasis, we also studied long-term responses to selection when epistatic interactions between beneficial alleles were strictly positive or negative, as opposed to an equal mixture of both. The results for positive epistasis were qualitatively similar to those for sign epistasis, with simulations of highly recombining species having stronger selection responses for the 10-locus trait (Figure S9 in File S1). However, in simulations with three loci, highly recombining bacteria showed similar or greater responses with positive or negative epistasis than those that recombine less, but the general difference between bacteria and eukaryotes remained. In the case of negative epistasis between 10 beneficial alleles, recombination generally produced greater selection responses, but these diminished as the cumulative effect of epistatic interactions approached that of additive effects (Ns_a = 10 per locus; Figure S9 in File S1), as the benefit of acquiring a new (and beneficial) allele was outweighed by the negative interactions it produced. When negative epistatic effects were much stronger than additive effects, the results began to resemble simulations with positive epistasis since the lack of a negative interaction between null alleles produced a benefit in effect, that far outweighed the deleterious epistatic effect of having both beneficial alleles. For the three-locus trait, selection responses for bacteria were greater than asexual simulations only for weak negative epistatic effects, but similar to the asexual for strong epistasis. These three locus dynamics of positive and negative epistasis appear, at least superficially, similar to two locus simulations of eukaryotes with different levels of recombination (Hansen 2013); simulations with a lower recombination rate had faster responses to selection with positive epistasis, but similar responses with negative epistasis since the effect of acquiring an additional beneficial allele via recombination diminishes with stronger negative epistatic interactions.

Distributions of pairwise LD between synonymous SNPs

Our finding that even high rates of bacterial recombination permit epistatic alleles to contribute to adaptation raises the question of whether epistasis has shaped patterns of PC, which we used to quantify recombination under a neutral model. While we attempted to mitigate the affect of selection on parameter estimates by using synonymous SNPs, these polymorphisms could theoretically exhibit epistatic interactions or be partially linked to epistatic nonsynonymous SNP pairs, both of which may skew distributions of LD (and thus PC; Kouyos *et al.* 2006) and lead to underestimates of population recombination rates if we assume these polymorphisms are neutral.

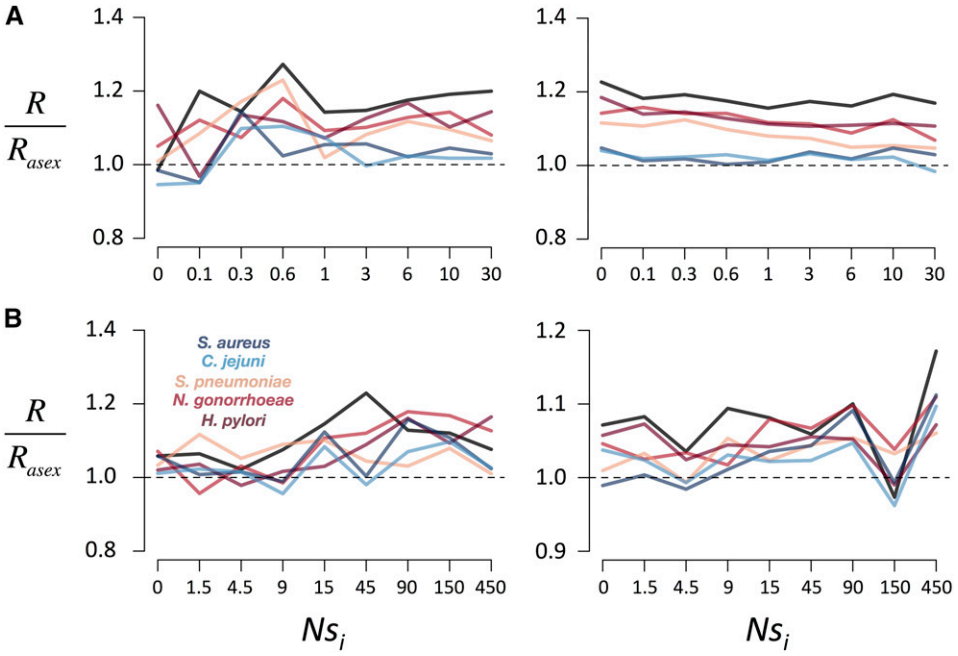

Figure 4 The effect of starting conditions on relative responses to selection. Shown are selection responses relative to an asexual control (R/R_{asex}) when recombination is suppressed during selection, illustrating how levels of haplotype diversity from mutation-recombination-drift balance affect selection responses, as opposed to the recombination that occurs during selection. Here, relative selection responses are measured after 0.2 N generations for a 10-locus trait (A) or a three-locus trait (B). For each plot, the mean of 2000 simulations is shown for each parameter set. Colors follow the same scheme as Figure 2. Epistatic effects per locus pair are shown on the x-axis, and additive effects per locus increase from the left column to the right column, with Ns_a = 1 or 10 for (A) or Ns_a = 3.33 or 33.3 for (B).

However, whether or not epistasis affects genome-wide patterns of LD depends on the frequency and strength of epistatic interactions, and whether selection is ongoing in the population (*i.e.*, loci under selection have not fixed). The full results, presented in the Supplemental Text and Table S5 in File S1, showed that distributions of LD were largely consistent with neutrality, although we cannot exclude selection having small effects on parameter estimation.

Discussion

Simple evolutionary models from population genetics shape our expectations and interpretations of nature and guide the development of future research (Fisher 1918; Wright 1931; Kimura 1965). Interest in sexual eukaryotes has inspired the simplifying convention of gene-centered models that promote the importance of additive gene effects and ignore epistasis; multilocus genotypes in eukaryotes rarely persist longer than a single generation such that only anomalously strong interactions require consideration (Crow 2001). This view pervades discussions of evolution and adaptation (Fisher 1918, 1937; Wright 1931; Coyne *et al.* 2000; Goodnight and Wade 2000), including those on bacteria, many of which recombine extensively and are commonly reduced to "core" and "accessory" genes, with niches providing selective advantages to specific genes (Fraser *et al.* 2007, 2009). While considering genes in isolation is certainly inappropriate for largely clonal bacteria, we find that this is may also be the case for highly recombining bacteria thought to be "effectively sexual" (Smith *et al.* 1993) from very low correlations between SNPs (r^2, Figure S10 in File S1). Short term adaptation proceeds rapidly even if the epistasis is weak ($N|s_i| \approx$ 3–10) and loci are distantly spaced, conditions that would make interactions virtually invisible to selection in eukaryotes (although this depends on the timescale under consideration; Paixão and Barton 2016), allowing bacteria to harness these effects to quickly respond to novel pressures and maintain beneficial allelic combinations in the face of extensive recombination (Cui *et al.* 2015; Skwark *et al.* 2017). Selection may act on even weaker epistatic effects for fitness traits controlled by >10 loci (the maximum explored here), since increasing the number of loci with fitness effects increases the total amount of selection on the trait, and, thus, relative amounts of recombination and selection (Neher and Shraiman 2009). The stark difference in the ability of epistatic alleles to contribute to short-term adaptation in bacteria and eukaryotes is driven by differences in genome architecture. Crossover recombination in eukaryotes exchanges large genomic segments, breaking numerous allele pairings even under low rates of exchange. Homologous recombination in bacteria, alternatively, involves the transfer of shorter DNA segments that can substantially reduce LD between nearly neutral polymorphisms over time (Figure S10 in File S1) but less so over the shorter evolutionary timescales on which selection acts and which we consider here. The differences we observe between bacterial and eukaryotic simulations are likely conservative since we only model a single chromosome, whereas multi-locus traits in eukaryotes are likely controlled by loci on different chromosomes that segregate randomly.

We have quantified bacterial recombination parameters with a new method sensitive enough to detect small tract lengths and events between closely related strains. We summarize these recombination parameters as F_{rec} (Table 2), which affects the selection response in the presence of epistasis in a complex way. While recombination breaks favorable allele combinations and generates recombination load, it also reduces clonal interference between beneficial mutations (or

Hill-Robertson interference; Hill and Robertson 1966) and allows greater exploration of fitness landscapes, which likely becomes more important as the number of loci (and thus possible allele combinations) increases and as the fitness landscape becomes increasingly rugged from sign epistasis. We observe that, for weak epistasis, recombination for the bacteria in this study is not strong enough to diminish short-term selection responses for either simple (three loci) or more complex (10 loci) traits when compared responses in an asexual organism. Hence the evolutionary dynamics of even the most highly recombining bacteria are more similar to fully asexual organisms rather than sexual eukaryotes. For strong epistasis, selection responses for bacteria become stronger than in asexuals likely due to the ability of recombination to reduce Hill-Robertson interference and allow more exploration of the fitness landscape, both during, but also before, selection if loci were previously under mutation-recombination-drift balance, as modeled here.

The countervailing effects of recombination on multilocus selection—breaking favorable allele combinations (hindering adaptation) or reducing Hill-Robertson interference (facilitating adaptation)—change in strength as the number of loci increases, with high bacterial recombination limiting short-term selection responses for simple traits (three loci) but accelerating selection for complex traits (10 loci). This trend is particularly evident for stronger epistasis ($N|s_i| > 10$) and depends on F_{rec}. As a comparison, we also show F_{rec} for the eukaryote modeled in this study, using a tract length of one-quarter the chromosome size C. While the actual tract length should be $C/2$, any exchanges B base pair longer than $C/2$ would break the same number of nucleotide pairs as $C - (C/2 + B)$. For example, an exchange of length $C/4$ would uncouple the same number of SNPs as one of length $3C/4$. Consequently, the most highly recombining bacteria modeled in this study would need to have an F_{rec} that is larger by more than an order of magnitude in order to exhibit eukaryote-like selection dynamics. Either increasing homologous recombination rates or tract lengths could achieve this.

The ability of bacterial recombination to potentially accelerate adaptation has been shown before for additive effects (Cohen *et al.* 2005; Cooper 2007; Levin and Cornejo 2009) and for epistatic fitness landscapes interspersed with fitness minima (Moradigaravand and Engelstädter 2012). Here, we use a general model of sign epistasis and explore a range of epistatic effects for simple and complex fitness landscapes to show how bacterial adaptation compares to asexuals, and the point at which recombination begins to accelerate selection responses in the presence of epistasis. Our results also show how recombination has distinct effects for simple and complex traits, and suggests that complex traits should be more accessible to highly recombining species. However, the ability of bacterial recombination to drive differences in selection dynamics between simple and complex traits diminishes if epistatic interactions are strictly positive or negative (Figure S9 in File S1).

These findings generate experimentally testable predictions relevant to exploring the distribution of epistatic effects between natural polymorphisms. For simple traits with sign epistasis, recombination largely antagonizes selection, such that effect sizes (in terms of Ns) may generally be stronger for species that recombine extensively; for instance compensatory mutations for antibiotic resistance may be stronger in species such as *N. gonorrheae* and *H. pylori*, or even less likely to exist, but this will also depend on starting allele frequencies as recombination has less of an effect when alleles are at extreme frequencies. This trend may reverse for more complex epistatic traits controlled by more loci, as the cumulative effect of weaker epistatic interactions produces greater selection responses for more highly recombining bacteria. Interestingly, our findings also show that moderate selection on a multi-locus trait may drive selection responses that do not lead to appreciable increases in pairwise LD between loci, since each epistatic locus pair contributes only a small effect. It may thus be difficult to detect weak (but important) epistasis in multilocus traits using pairwise LD measures from population genomic data.

We primarily use sign epistasis in our simulations (*Materials and Methods*), which was used in historical arguments to challenge Fisher on the potential importance of gene interactions (Wright 1931; Wade 1992) but has also been frequently observed in empirical fitness landscapes (Weinreich *et al.* 2005; Weinreich 2006; Poelwijk *et al.* 2007; Kvitek and Sherlock 2011; Silva *et al.* 2011). While the specific type of epistasis will have important consequences for long-term evolution, such as the evolution of recombination (Kondrashov 1988; Barton 1995b), that is not our focus here. We have focused on the short term to define the fundamental capacity of selection to act on allele combinations in the face of bacterial recombination. Ultimately, the importance of epistasis will also depend on the distribution of epistatic fitness effects, which may vary between traits, and the number of loci that contribute to multi-locus fitness traits in real populations.

It is important to note that we initiate simulations of selection with polymorphisms from mutation-recombination-drift balance, essentially studying how neutral-equilibrium levels of LD and recombination interact to affect selection dynamics. Our results would likely be different if there was also epistasis between these polymorphisms selected from equilibrium conditions; positive LD between polymorphisms would accumulate from selection on positive epistatic interactions,

Table 2 Fraction of bases exchanged per generation via homologous recombination (F_{rec})

Species	$\rho/q(Chromosome\ Size)$[a]
S. aureus	3.91×10^{-7}
C. jejuni	4.29×10^{-7}
S. pneumoniae	3.38×10^{-6}
N. gonorrheae	1.08×10^{-5}
H. pylori	1.17×10^{-5}
Yeast	5.0×10^{-4}

[a] Recombination parameter values from MDEs in Table 1, chromosome size used was 2 Mb.

and negative LD from negative interactions (Eshel and Feldman 1970). Since recombination only increases the additive variance in fitness when it uncouples polymorphisms in negative LD (Charlesworth 1993; Barton 1995; but see Barton (2010) for a review), it would have a much larger affect on selection dynamics if mutations started out with more negative associations. Likewise, if our simulations of selection were initiated with polymorphisms that were previously deleterious (*i.e.*, negative additive effects), this would push allele frequencies toward more extreme values (near 0 or 1; Charlesworth *et al.* 1993; Kim 2006) and generate more negative LD since selection would more quickly eliminate haplotypes with more deleterious mutations (Barton and Otto 2005). According to eukaryotic models, shifting alleles toward more extreme frequencies would increase the additive genetic variance (and decrease the epistatic variance; Hill *et al.* 2008) and diminish the impact of recombination on short-term dynamics, since recombination is less likely to generate new haplotypes or break existing ones. However, recombination would still likely have important long-term effects as selection on epistatic interactions changes allele frequencies and haplotype diversity over time (Paixão and Barton 2016), and, thus, the probability that recombination generates novel diversity and facilitates exploration of fitness landscapes.

While our dataset consists of opportunistic and obligate pathogens, our results likely extend to other microbes. For instance, a genomic study of *Vibrio cyclitrophicus* showed recombination is sufficiently strong to allow gene-specific selective sweeps, as opposed to the periodic selection model in which sweeps have genome-wide effects (Shapiro *et al.* 2012). Like in *H. pylori*, LD decays rapidly within 50 bp in *V. cyclitrophicus*, but asymptotes at a value well above zero, suggesting residual linkage genome-wide. Thus, while gene flow and recombination may homogenize ecotypes outside of genomic regions involved in local adaptation, recombination may not be strong enough to antagonize selection on epistatic interactions. However, exact estimates of recombination from LD data require knowledge of population size (via knowledge of the mutation rate). We also applied our method to a genomic dataset of thermophilic archaea *Sulfolobus islandicus* (Cadillo-quiroz *et al.* 2012) but struggled to accurately infer the mean tract length due to low diversity and small sample size (11 isolates). Nevertheless, our best parameter estimates strongly suggest a low recombination rate that likely permits selection on weak epistatic interactions (Figure S11 in File S1). Thus, epistasis between polymorphisms scattered across the genome may play a critical role in adaptation for the majority of the tree of life, and unlike eukaryotes, these interactions do not need to be strong or physically close, and do not require specific metapopulation dynamics to permit efficient selection.

Acknowledgments

All data used in this study came from previously published papers shown in Table S1 in File S1. B.J.A. was supported by a postdoctoral fellowship F32 GM120839-01, M.L. by research grant R01AI048935, and W.P.H. by R01AI106786 from the National Institutes of Health (www.nih.gov). These authors were additionally supported by the MIDAS program (U54GM088558). The computations in this paper were run on the Odyssey cluster supported by the Faculty of Arts and *Sciences* (FAS) Division of Science, Research Computing Group at Harvard University. The authors declare no competing financial interests.

Author contributions: B.J.A., W.P.H., and M.L. conceived and headed the project. B.J.A., M.U.G., and J.C. designed the statistical framework used to infer recombination parameters. B.J.A. performed all simulations and bioinformatic analyses. B.J.A., W.P.H., and M.L. wrote the manuscript with input from all coauthors. S.K.S. provided data for *Campylobacter jejuni*, and Y.H.G. provided data and expertise for *Neisseria gonorrhoeae*.

Literature Cited

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

Barton, N. H., 1995a Linkage and the limits to natural selection. Genetics 140: 821–841.

Barton, N. H., 1995b A general model for the evolution of recombination. Genet. Res. 65: 123–145.

Barton, N. H., 2010 Genetic linkage and natural selection. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365: 2559–2569.

Barton, N. H., and S. P. Otto, 2005 Evolution of recombination due to random drift. Genetics 169: 2353–2370.

Barton, N. H., and M. Turelli, 2004 Effects of genetic drift on variance components under a general model of epistasis. Evolution 58: 2111–2132.

Bubendorfer, S., J. Krebes, I. Yang, E. Hage, T. F. Schulz *et al.*, 2016 Genome-wide analysis of chromosomal import patterns after natural transformation of Helicobacter pylori. Nat. Commun. 7: 11995.

Bulmer, M. G., 1976 The effect of selection on genetic variability: a simulation study. Genet. Res. 28: 101–117.

Bulmer, M. G., 1980 *The Mathematical Theory of Quantitative Genetics*. Oxford University Press, Oxford.

Cadillo-quiroz, H., X. Didelot, N. L. Held, A. Herrera, A. Darling *et al.*, 2012 Patterns of gene flow define species of thermophilic archaea. PLoS Biol. 10: e1001265.

Carter, A. J. R., J. Hermisson, and T. F. Hansen, 2005 The role of epistatic gene interactions in the response to selection and the evolution of evolvability. Theor. Popul. Biol. 68: 179–196.

Charlesworth, B., 1993 Directional selection and the evolution of sex and recombination. Genet. Res. 61: 205–224.

Charlesworth, B., M. T. Morgan, and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics 134: 1289–1303.

Cheverud, J. M., and E. J. Routman, 1996 Epistasis as a source of increased additive genetic variance at population bottlenecks. Evolution 50: 1042–1051.

Cohen, E., D. A. Kessler, and H. Levine, 2005 Recombination dramatically speeds up evolution of finite populations. Phys. Rev. Lett. 94: 098102.

Cooper, T. F., 2007 Recombination speeds adaptation by reducing competition between beneficial mutations in populations of Escherichia coli. PLoS Biol. 5: e225.

Coyne, J., N. Barton, and M. Turelli, 2000 Is Wright's shifting balance process important in evolution? Evolution 54: 306–317.

Croucher, N. J., A. J. Page, T. R. Connor, A. J. Delaney, J. A. Keane *et al.*, 2014 Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 43: e15.

Crow, J. F., 2001 The beanbag lives on. Nature 409: 771.

Crow, J. F., 2010 On epistasis: why it is unimportant in polygenic directional selection. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365: 1241–1244.

Cui, Y., X. Yang, X. Didelot, C. Guo, D. Li *et al.*, 2015 Epidemic clones, oceanic gene pools, and eco-LD in the free living marine pathogen vibrio parahaemolyticus. Mol. Biol. Evol. 32: 1396–1410.

Darling, A. C. E., B. Mau, F. R. Blattner, and N. T. Perna, 2004 Mauve : multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14: 1394–1403.

De Maio, N., and D. J. Wilson, 2017 The bacterial sequential Markov coalescent. Genetics 206: 333–343. https://doi.org/10.1534/genetics.116.198796.

Eshel, I., and M. W. Feldman, 1970 On the evolutionary effect of recombination. Theor. Popul. Biol. 1: 88–100.

Falush, D., C. Kraft, N. S. Taylor, P. Correa, J. G. Fox *et al.*, 2001 Recombination and mutation during long-term gastric colonization by Helicobacter pylori: estimates of clock rates, recombination size, and minimal age. Proc. Natl. Acad. Sci. USA 98: 15056–15061.

Fisher, R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edinb. 52: 399–433.

Fisher, R. A., 1930 *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Fisher, R. A., 1937 *The Design of Experiments*. Oliver & Boyd, Edinburgh.

Fraser, C., W. P. Hanage, and B. G. Spratt, 2007 Recombination and the nature of bacterial speciation. Science 315: 476–480.

Fraser, C., E. J. Alm, M. F. Polz, B. G. Spratt, and W. P. Hanage, 2009 The bacterial species challenge: making sense of genetic and ecological diversity. Science 323: 741–746.

Goodnight, C. J., 1988 Epistasis and the effect of founder events on the additive genetic variance. Evolution 42: 441–454.

Goodnight, C. J., and M. J. Wade, 2000 The ongoing synthesis: a reply to Coyne, Barton, and Turelli. Evolution 54: 317–324.

Gupta, S., M. C. Maiden, I. M. Feavers, S. Nee, R. M. May *et al.*, 1996 The maintenance of strain structure in populations of recombining infectious agents. Nat. Med. 2: 437–442.

Gutmann, M. U., and J. Corander, 2016 Bayesian optimization for likelihood-free inference of simulator-based statistical models. J. Mach. Learn. Res. 17: 1–47.

Hallander, J., and P. Waldmann, 2007 The effect of non-additive genetic interactions on selection in multi-locus genetic models. Heredity (Edinb) 98: 349–359.

Hansen, T. F., 2013 Why epistasis is important for selection and adaptation. Evolution 67: 3501–3511.

He, X., W. Qian, Z. Wang, Y. Li, and J. Zhang, 2010 Prevalent positive epistasis in Escherichia coli and Saccharomyces cerevisiae metabolic networks. Nat. Genet. 42: 272–276.

Hernandez, R. D., 2008 A flexible forward simulator for populations subject to selection and demography. Bioinformatics 24: 2786–2787.

Hill, W. G., and A. Mäki-Tanila, 2015 Expected influence of linkage disequilibrium on genetic variance caused by dominance and epistasis on quantitative traits. J. Anim. Breed. Genet. 132: 176–186. https://doi.org/10.1111/jbg.12140.

Hill, W. G., and A. Robertson, 1966 The effect of linkage on limits to artificial selection. Genet. Res. 52: 349–363.

Hill, W. G., M. E. Goddard, and P. M. Visscher, 2008 Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 4: e1000008.

Hoggart, C. J., M. Chadeau-Hyam, T. G. Clark, R. Lampariello, J. C. Whittaker *et al.*, 2007 Sequence-level population simulations over large genomic regions. Genetics 177: 1725–1731.

Hudson, R. R., 2002 Generating samples under a Wright–Fisher neutral model of genetic variation. Bioinforma. Appl. Note 18: 337–338.

Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111: 147–164.

Kennemann, L., X. Didelot, T. Aebischer, S. Kuhn, B. Drescher *et al.*, 2011 Helicobacter pylori genome evolution during human infection. Proc. Natl. Acad. Sci. USA 108: 5033–5038.

Kim, Y., 2006 Allele frequency distribution under recurrent selective sweeps. Genetics 172: 1967–1978.

Kimura, M., 1965 Attainment of quasi linkage equilibrium when gene frequencies are changing by natural selection. Genetics 52: 875–890.

Kondrashov, A. S., 1988 Deleterious mutations and the evolution of sexual reproduction. Nature 336: 435–440.

Kouyos, R. D., S. P. Otto, and S. Bonhoeffer, 2006 Effect of varying epistasis on the evolution of recombination. Genetics 173: 589–597.

Kvitek, D. J., and G. Sherlock, 2011 Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. PLoS Genet. 7: e1002056.

Levin, B. R., and C. T. Bergstrom, 2000 Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. Proc. Natl. Acad. Sci. USA 97: 6981–6985.

Levin, B. R., and O. E. Cornejo, 2009 The population and evolutionary dynamics of homologous gene recombination in bacterial populations. PLoS Genet. 5: e1000601.

Lin, M., and E. Kussell, 2016 Correlated mutations and homologous recombination within bacterial populations Mingzhi. Genetics 205: 891–917.

Mailund, T., M. H. Schierup, C. N. S. Pedersen, P. J. M. Mechlenborg, J. N. Madsen *et al.*, 2005 CoaSim: a flexible environment for simulating genetic data under coalescent models. BMC Bioinformatics 6: 252.

Mäki-Tanila, A., and W. Hill, 2014 Influence of gene interaction on complex trait variation with multilocus models. Genetics 198: 355–367.

Martin, G., S. F. Elena, and T. Lenormand, 2007 Distributions of epistasis in microbes fit predictions from a fitness landscape model. Nat. Genet. 39: 555–560.

Méric, G., M. Miragaia, M. de Been, K. Yahara, B. Pascoe *et al.*, 2015 Ecological overlap and horizontal gene transfer in staphylococcus aureus and staphylococcus epidermidis. Genome Biol. Evol. 7: 1313–1328.

Moradigaravand, D., and J. Engelstädter, 2012 The effect of bacterial recombination on adaptation on fitness landscapes with limited peak accessibility. PLoS Comput. Biol. 8: e1002735.

Neher, R. A., and B. I. Shraiman, 2009 Competition between recombination and epistasis can cause a transition from allele to genotype selection. Proc. Natl. Acad. Sci. USA 106: 6866–6871.

Ohta, T., 1976 Role of very slightly deleterious mutations in molecular evolution and polymorphism. Theor. Popul. Biol. 10: 254–275.

Page, A. J., C. A. Cummins, M. Hunt, V. K. Wong, S. Reuter *et al.*, 2015 Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31: 3691–3693.

Paixão, T., and N. H. Barton, 2016 The effect of gene interactions on the long-term response to selection. Proc. Natl. Acad. Sci. USA 113: 4422–4427.

Poelwijk, F. J., D. J. Kiviet, D. M. Weinreich, and S. J. Tans, 2007 Empirical fitness landscapes reveal accessible evolutionary paths. Nature 445: 383–386.
Roychoudhury, A., and J. Wakeley, 2010 Sufficiency of the number of segregating sites in the limit under finite-sites mutation. Theor. Popul. Biol. 78: 118–122.
Seemann, T., 2014 Prokka: rapid prokaryotic genome annotation. Bioinformatics 30: 2068–2069.
Shapiro, B. J., J. Friedman, O. X. Cordero, S. P. Preheim, S. C. Timberlake *et al.*, 2012 Population genomics of early events in the ecological differentiation of bacteria. Science 336: 48–51.
Silva, R. F., S. C. M. Mendonça, L. M. Carvalho, A. M. Reis, I. Gordo *et al.*, 2011 Pervasive sign epistasis between conjugative plasmids and drug-resistance chromosomal mutations. PLoS Genet. 7: e1002181.
Skwark, M. J., N. J. Croucher, S. Puranen, C. Chewapreecha, M. Pesonen *et al.*, 2017 Interacting networks of resistance, virulence and core machinery genes identified by genome-wide epistasis analysis. PLoS Genet. 13: e1006508.
Slatkin, M., 1994 Linkage disequilibrium in growing and stable populations. Genetics 137: 331–336.
Slatkin, M., 2008 Linkage disequilibrium–understanding the evolutionary past and mapping the medical future. Nat. Rev. Genet. 9: 477–485.
Smith, J. M., N. H. Smith, M. O'Rourke, and B. G. Spratt, 1993 How clonal are bacteria? Proc. Natl. Acad. Sci. USA 90: 4384–4388.
Suerbaum, S., J. M. Smith, K. Bapumia, G. Morelli, N. H. Smith *et al.*, 1998 Free recombination within Helicobacter pylori. Proc. Natl. Acad. Sci. USA 95: 12619–12624.
Tajima, F., 1996 The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. Genetics 143: 1457–1465.
Trindade, S., A. Sousa, K. B. Xavier, F. Dionisio, M. G. Ferreira *et al.*, 2009 Positive epistasis drives the acquisition of multidrug resistance. PLoS Genet. 5: e1000578.
Wade, M. J., 1992 Sewall Wright, gene interaction and the shifting balance theory, pp. 35–62 in *Oxford Surveys in Evolutionary Biology*, Vol. 8, edited by D. Futuyma, and J. Antonovics. Oxford University Press, New York.
Weinreich, D. M., 2006 Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312: 111–114. https://doi.org/10.1126/science.1123539.
Weinreich, D. M., R. A. Watson, and L. Chao, 2005 Perspective: sign epistasis and genetic constraint on evolutionary trajectories. Evolution 59: 1165–1174.
Whitlock, M. C., P. C. Phillips, F. B. G. Moore, and S. J. Tonsor, 1995 Multiple fitness peaks and epistasis. Annu. Rev. Ecol. Syst. 26: 601–629.
Wilson, E. O., 1965 A consistency test for phylogenies based on contemporaneous species. Syst. Zool. 14: 214–220.
Wright, S., 1931 Evolution in Mendelian populations. Genetics 16: 97–159.
Yukilevich, R., J. Lachance, F. Aoki, and J. R. True, 2008 Long-term adaptation of epistatic genetic networks. Evolution 62: 2215–2235 (erratum: Evolution 62: 2951).

Communicating editor: L. Wahl