nature ecology & evolution

ARTICLES

PUBLISHED: 21 NOVEMBER 2016 | VOLUME: 1 | ARTICLE NUMBER: 0007

Competition along trajectories governs adaptation rates towards antimicrobial resistance

C. Brandon Ogbunugafor[1,2†] and Margaret J. Eppstein[2,3*†]

The increasing availability of genotype–phenotype maps for different combinations of mutations has empowered evolutionary biologists with the tools to interrogate the predictability of adaptive evolution, especially in the context of the evolution of antimicrobial resistance. Large microbial populations are known to generate competing beneficial mutations, but determining how these mutations contribute to the adaptive trajectories that are most likely to be followed remains a challenge. Despite a recognition that there may also be competition between successive alleles on the same trajectory, prior studies have not fully considered how this impacts adaptation rates along, or likelihood of following, individual trajectories. Here, we develop a metric that quantifies the competition between successive alleles along adaptive trajectories and show how this competition largely governs the rate of evolution in simulations on empirical fitness landscapes for proteins involved in drug resistance in two species of malaria (*Plasmodium falciparum* and *P. vivax*). Our findings reveal that a trajectory with a larger-than-average initial fitness increase may have smaller fitness increases in later steps, which slows adaptation. In some circumstances, these trajectories may be outcompeted by alleles on faster alternative trajectories that are being explored simultaneously. The ability to predict adaptation rates along accessible trajectories has implications for efforts to manage antimicrobial resistance in real-world settings and for the broader intellectual pursuit of predictive evolution in complex adaptive fitness landscapes for a variety of application domains.

Since Sewall Wright first introduced the landscape analogy to conceptualize the relationship between genotypes and reproductive fitness[1] it has proved to be a useful theoretical framework not only in biological systems but also in complex technological[2], computational[3], and social systems[4]. This analogy is widely used in evolutionary biology, with a growing number of genotype–phenotype maps generated from empirical data contributing to foundational studies of drug resistance in several microbial systems[5,6]. The increasing availability of these empirical fitness landscapes enables us to probe to what degree evolution is predictable[7,8] and could potentially improve treatment regimens[9]. It is becoming clear that sign epistasis is ubiquitous in many of these systems[10,11] and that the resulting landscape 'ruggedness' may constrain evolution to a relatively small set of accessible trajectories of increasing fitness[5,6,8,12–14]. Yet recent experimental studies of rugged empirical fitness landscapes[15,16] underscore that we have yet to fully resolve all of the essential properties that dictate why certain trajectories are traversed. If one invokes the assumptions of the strong selection and weak mutation regime (SSWM), then fixation of a new beneficial mutation is assumed to occur essentially instantaneously[7], and it has been assumed that the likelihood of fixation of an allele is correlated with its selective advantage[5,6]. However, given the observed population sizes and mutation rates for many microbial systems[17–20], it is likely that multiple alleles may coexist in competition with each other for some time, thus violating the assumptions of the SSWM regime[15,21,22]. Previous work on such clonal interference has shown that the likelihood that evolution will favour the so-called 'greedy' trajectory (the one that selects the available allele with the largest fitness increase at each step) actually decreases at these large population sizes[23,24].

We distinguish between two types of clonal interference. 'Between-path' clonal interference refers to competition between alleles on different trajectories (Fig. 1), and has previously received considerable attention[16,23,24]. In contrast, 'within-path' clonal interference refers to competition between successive alleles on the same adaptive trajectory (Fig. 1). While it is recognized that the time to fixation of a new beneficial allele is a function of the amount of within-path competition with its predecessor[22], we have found no prior work that quantifies the effect of within-path clonal interference on the adaptation time along a multi-allele trajectory or examines the impact that this may have on the likelihood that the trajectory will be followed.

Results

We explored adaptation rates and the trajectories taken using a discrete population model to simulate evolution in asexually reproducing populations of microorganisms such as malaria in the host (see Methods). We represent bi-allelic haploid genotypes as binary strings, where 0 or 1 mean wild-type or mutant loci, respectively. Binary strings that differ at exactly one locus are referred to as 'mutational neighbours'. We examined 29 different adaptive landscapes for two sets of four bi-allelic loci in the dihydrofolate reductase (DHFR) gene that are known to be associated with antimicrobial resistance in malaria. We studied 19 adaptive landscapes for *P. falciparum*, corresponding to one drug-free landscape, nine concentrations of pyrimethamine (Supplementary Table 1), and nine concentrations of cycloguanil (Supplementary Table 2). Ten landscapes were studied for *P. vivax*, corresponding to one drug-free landscape and nine concentrations of pyrimethamine (Supplementary Table 3). We ran simulations starting from each of the 16 genotypes in each landscape (referred to as 'seed' genotypes), using realistically large population sizes and low mutation rates (see Methods). Of the 464 unique landscape–seed combinations, 20 were seeded with inviable genotypes and 29 were seeded with the global optimum, and so did not require simulation. Of the remaining 415 cases, 60 always became indefinitely trapped on a local optimum, 3 sometimes became trapped

[1]Department of Biology, University of Vermont, Burlington, Vermont 05405, USA. [2]Vermont Complex Systems Center, University of Vermont, Burlington, Vermont 05405, USA. [3]Department of Computer Science, University of Vermont, Burlington, Vermont 05405, USA. [†]These authors contributed equally to this work. *e-mail: maggie.eppstein@uvm.edu

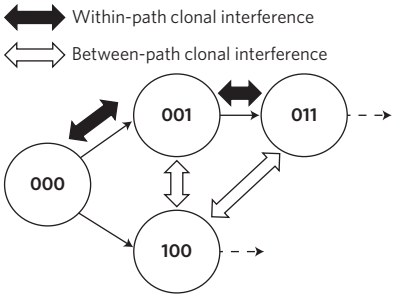

Figure 1 | Types of clonal interference. Consider a hypothetical evolutionary trajectory (000 → 001 → 011 → …, where 0 or 1 represent wild-type or mutant loci, respectively) with monotonically increasing fitness. The competition between each allele and its ancestor on the same trajectory (filled double-arrows) is what we term within-path clonal interference. Competition with co-existing alleles on alternate trajectories (for example, 100) is what we term between-path clonal interference (open double-arrows).

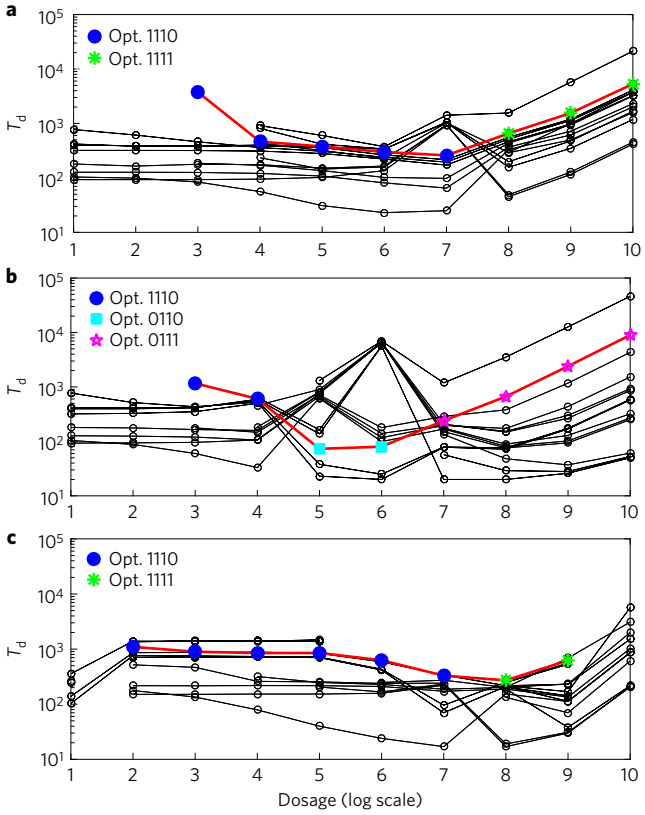

Figure 2 | Speed of acquisition of drug resistance varies non-monotonically with dosage. The time it takes for the global optimum to become dominant (T_d) varies non-monotonically with dosage on the 29 empirical fitness landscapes tested. Each line represents simulations starting from a different seed genotype, with the red line corresponding to simulations that start from the wild-type and reach the dosage-specific global optima indicated by the coloured symbols. Dosage levels correspond to the columns in Supplementary Tables 1–3.

and sometimes evolved to the global optimum and 352 consistently evolved to became dominated by the global optimum. (As with any study of empirical fitness landscapes, these conclusions are necessarily limited by the extent of the mapped landscapes.)

We observed large disparities in the time it takes for the optimal genotype to become dominant (T_d) along different evolutionary trajectories, even for trajectories of the same length. Interestingly, for a given combination of the species of malaria, drug, seed genotype and global optimum, the time it takes to evolve maximum resistance varies non-monotonically with dosage (Fig. 2). Closer examination revealed that the time it takes for a new beneficial mutation to outcompete its immediate predecessor is governed by the reciprocal of the difference between their growth rates (that is, the fitness gradient). The steeper the gradient, the more rapidly a beneficial mutation will become fixed. Thus for an accessible evolutionary trajectory with strictly increasing fitnesses, we quantify the total within-path competition (C_w) along the trajectory as the sum of the reciprocal differences in the growth rates between all pairs of adjacent genotypes along the path, as illustrated in Fig. 3a (see Methods for a derivation of this metric). The reciprocal of the smallest growth rate differences (the flattest gradients) on a trajectory therefore dominate the value of C_w. One can think of the sum of reciprocal growth rate gradients as a measure of resistance to movement along the trajectory, much as the total resistance in a serial electrical circuit is the sum of all of the resistors along the circuit. The occurrence of a larger-than-average initial fitness increase (as often occurs on the greediest path) necessitates smaller fitness increases further down the path towards a given optimum, thus increasing C_w and T_d (Fig. 3b,c; blue lines) relative to a trajectory with equal step sizes (the latter being the fastest possible trajectory) (Fig. 3b,c; green lines). The surprisingly simple and intuitive metric C_w allows the comparison of within-path competition along different adaptive trajectories (on the same or different landscapes), providing insight into the reason for wide disparities in adaptation rates along different trajectories, and revealing why greedy trajectories are not always the most likely to be followed.

If the initial fitness increase in the greediest trajectory is sufficiently higher than the fitness of competing alleles, the greediest trajectory will outcompete faster alternative trajectories on the same landscape. For example, if the two trajectories shown in Fig. 3b,c are explored simultaneously, the greedy (blue) trajectory always wins, because the second node on the greedy trajectory outcompetes the second node on the faster (green) trajectory before the latter has a chance to mutate further. However, if the faster trajectory is only slightly less greedy, it may win the competition. The hypothetical paths shown in Fig. 3d illustrate how this can happen. Although the path shown in blue is initially slightly greedier than the green path, the subtle differences in the growth rates translate into large differences in C_w. Simulations along each of these trajectories independently (Fig. 3e) confirm that the greedier trajectory is much slower. In simulations where they are allowed to compete, the faster trajectory consistently beats the greedier trajectory.

For example, consider the competition between the two trajectories that are simultaneously being explored on the empirical fitness landscape illustrated in Fig. 4a,b. Genotype 0110 on the slightly less greedy but faster trajectory (dashed lines, $C_w = 62.1$) outcompeted genotype 1001 on the greediest trajectory (solid lines, $C_w = 84.3$) in 1,000 out of 1,000 simulations. The fact that, in this case, the greedy trajectory leads to a different peak (which is slightly lower, although only by an amount that is probably less than experimental error) is irrelevant, since the greedy path was consistently outcompeted before it ever even reached the peak 1101 (Fig. 4b). Once we had determined which trajectory first reached the global optimum (the 'winning' trajectory) in such simulations, we subsequently ran 'path-only' simulations, in which all between-path competition (C_b) was temporarily suppressed by setting all of the growth rates outside the winning trajectory to zero (for example, Fig. 4c,d). This enabled us to separately quantify the effects of C_w and C_b on adaptation rates. Although the fitness differences between the greedy and less-greedy paths shown in Fig. 4a,b are relatively small, the same thing can

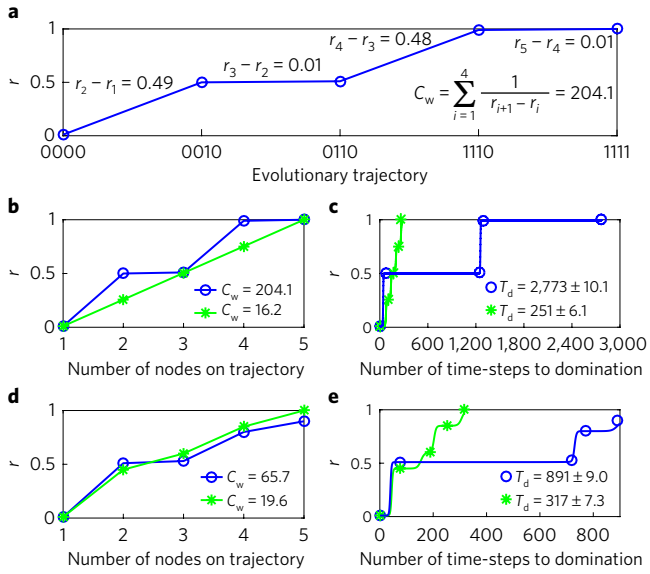

Figure 3 | Quantifying within-path competition. a, C_w along a hypothetical evolutionary trajectory with four steps (0000 → 0010 → 0110 → 1110 → 1111) is quantified as the sum of the reciprocal differences in growth rates ($r_{i+1} - r_i$) between adjacent genotypes i and $i+1$ on the trajectory (see equation (16)). **b**, Two different hypothetical trajectories leading to the global optimum. C_w is minimized along the lower (green asterisks) trajectory where the changes in adjacent growth rates are equal. The larger initial fitness increase on the greedy path (blue open circles) means that one or more subsequent steps are necessarily smaller, thus increasing C_w, which is dominated by the smallest step size. **c**, Independent simulations along each of the two trajectories shown in **b** (not in competition with each other) reveal that the blue trajectory is much slower than the green trajectory (two-sided Wilcoxon rank sum, $P < 3.5 \times 10^{-8}$). **d**,**e**, Minor differences in $r_{i+1} - r_i$ along two hypothetical trajectories translate into large differences in C_w and resulting adaptation rates. The blue trajectory is only slightly greedier, but the green trajectory has much less within-path clonal interference and is thus much faster (two-sided Wilcoxon rank sum, $P < 3.5 \times 10^{-8}$). In **c** and **e** the x coordinate of each symbol represents the median number of time-steps (from $n = 21$ simulations) at which the maximum frequency of each of the intermediate alleles was achieved or the final allele became dominant, with the median ± the standard deviation shown in the legend. The lines show the average growth rate over the entire population for one representative simulation on each trajectory.

occur (albeit less frequently) with relatively large fitness differences (Supplementary Table 4, Supplementary Fig. 1).

The results of our simulations confirm that adaptation times in all 29 empirical fitness landscapes studied are largely governed by within-path competition along the winning trajectories (coefficient of determination $R^2 > 0.996$, Supplementary Fig. 2). When we subsequently ran the path-only simulations on these same trajectories, the strength of the relationship improved to $R^2 > 0.998$ (Fig. 5). Note that this relationship is equally strong if one only considers the 21 simulations from the non-overlapping trajectories where the seed is exactly 4 steps from the optimum (Supplementary Fig. 3). The path-only simulations reveal that much of the residual error in the regressions on the data from the full landscapes (Supplementary Fig. 2) can be explained by the existence of C_b due to clonal interference from genotypes on competing trajectories. The effects of such between-path clonal interference that occur at the final step become more pronounced when simulations are run until T_f, the time-step when the optimal genotype becomes fixed (exceeds 99% of the population, Supplementary Fig. 4). See Supplementary Fig. 5 for an illustration and discussion of additional sources of error that account for the remaining small residuals. The difference between the observed T_d or T_f on the empirical fitness landscapes and the path-only T_d or T_f observations, as indicated by the lengths of the red vertical lines in Fig. 5 and Supplementary Fig. 4, can be used to quantify C_b. Unlike C_w, this quantification of C_b requires simulations and is thus sensitive to the necessarily arbitrary nature of these stopping criteria.

The strength of the dependence of $\ln(T_d)$ on $\ln(C_w)$ in path-only simulations is insensitive to specific assumptions on the simulation parameters in realistic ranges, such as the maximum population size, mutation rate or absolute growth rates (Supplementary Table 5), although the specific coefficients are obviously affected by these parameters. For example, as the number of mutations increases (due to increases in any of these parameters), the relationship becomes increasingly sublinear (Supplementary Table 5) because ongoing mutational events increase the speed along the trajectory. Increasing stochasticity by using Poisson-distributed mutation rates (see Methods) causes the distributions to become increasingly negatively skewed below the best fit line shown in Fig. 5 because of the increasing occurrence of rare 'lucky' mutations that decrease convergence times (Supplementary Fig. 6).

We further examined the 242 non-trivial landscape–seed combinations that always or sometimes converged to the global optimum, and in which the seed differed from the global optimum in at least two loci. As expected, most of them (220) were found to follow the greediest trajectory from the seed to the optimum in 100 out of 100 trials. However, as found elsewhere[16,21,22], our simulations confirmed that multiple accessible pathways may be explored simultaneously and thus compete with each other. Indeed, in the empirical fitness landscapes we explored, we found 22 landscape–seed combinations where faster trajectories outcompeted the greedy trajectory some or all of the time. In 12 of these cases, the greedy trajectory happened to lead to a different peak, but the faster trajectory was able to win the competition and reach the global optimum (Supplementary Table 4; see Fig. 4a,b and Supplementary Fig. 1 for representative simulations). In the other 10 cases, greedy and slightly less-greedy trajectories of the same length were simultaneously followed towards the same global optimum (see Supplementary Fig. 7 for a representative illustration), with the less-greedy but faster trajectories generating the first optimal mutant in 2.4–46.5% of the trials (Supplementary Table 6). In 1,000 repetitions of each of these we found that: (1) T_d was significantly lower on the less greedy path in 9 of the 10 cases (2-sided Wilcoxon rank sum, $P \ll 10^4$, Supplementary Table 6); and (2) the proportion of the 1,000 repetitions in which the first individual of the optimal genotype came from the less greedy path was positively correlated with the amount by which it was faster than the greedy path (Pearson correlation $r = 0.69$, $P < 0.02$).

Discussion

It has previously been shown that small populations, or spatially structured populations with limited mixing, can explore alternative trajectories more effectively on rugged landscapes than large well-mixed populations. This occurs because the effects of genetic drift are amplified when selection pressure is reduced, thus enabling these populations to sometimes avoid becoming trapped on sub-optimal peaks[23,25–27]. Here we have shown that even when the greediest path leads to a suboptimal peak, very large populations can also sometimes reach a higher peak along a faster competing trajectory. Similarly, when competing paths lead to the same fitness peak, the faster but less-greedy path can sometimes get there first. In either case, it is not the fitness of the peak that is the determining factor as to which trajectory wins the competition. Rather, it is the amount of within-path competition that governs the evolutionary speed along these different trajectories, potentially enabling a less greedy trajectory to reach a higher fitness more quickly and therefore ultimately outcompete the slower but greedier trajectory.

On the *P. falciparum* DHFR landscapes that we studied for pyrimethamine dosages between 0.1 μM and 100 μM, the greedy

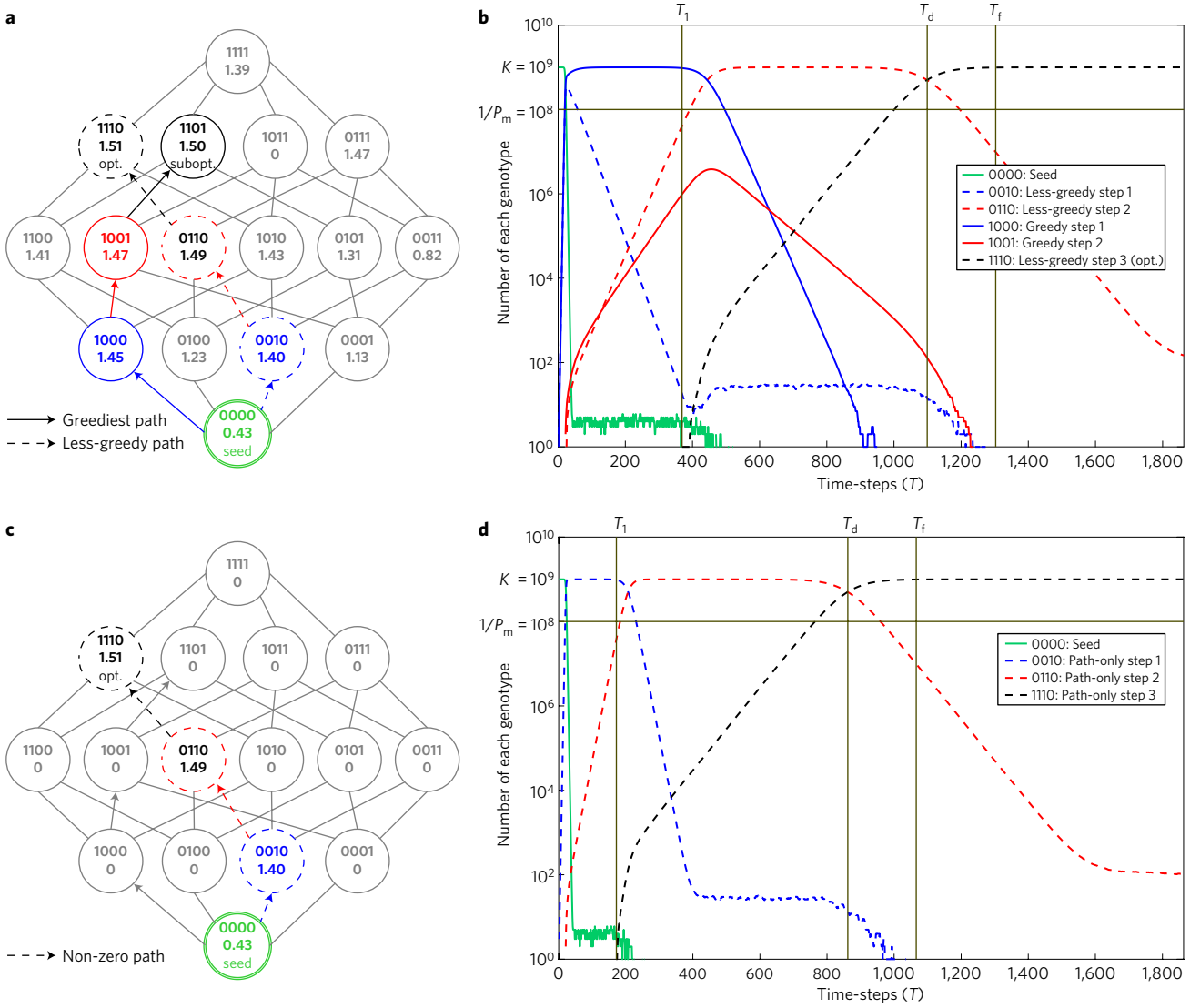

Figure 4 | Illustrative example of the influence of clonal interference on trajectories taken and adaptation rates. a, Hypercube depiction of the empirical fitness landscape for *P. vivax* treated with of 1.7 μM pyrimethamine, showing two competing trajectories starting from the wild-type genotype 0000. Genotypes and normalized growth rates are shown inside each node, with edges representing mutational neighbours in the landscape. **b**, Representative simulation illustrating competition between the two paths shown in **a**, showing that the slightly less-greedy but faster path (dashed lines, $C_w = 62.1$) outcompetes the greediest path (solid lines, $C_w = 84.3$). For visual clarity, only those genotypes on these two paths are shown. The *y* axis is shown on a log scale so that lower frequency genotypes can be observed; T_1, T_d and T_f refer to the generation in which the optimal genotype first appeared, became dominant and became fixed (>99%), respectively. **c**, A landscape in which we have set all growth rates to 0, except those on the path that reached the global optimum in the full landscape simulation shown in **b**. **d**, A representative path-only simulation on the landscape shown in **c**. Convergence times in such path-only simulations are governed by C_w. The difference in the convergence times between the full landscape simulations (for example, **b**) and the path-only simulations (for example, **d**) can be attributed to non-zero C_b in the full landscape simulations.

trajectory is 0000 → 0010 → 0110 → 1110. It is thus not surprising that all of these alleles have been observed in real-world settings[6]. However, a recent study of changes in pyrimethamine-resistant DHFR mutants in *P. falciparum* in the Gambia[28] found increases in mutants along a less-greedy trajectory with lower within-path competition (0000 → 0010 → 1010 → 1110). This provides intriguing (albeit indirect) evidence that evolution may have preferred the faster but less-greedy path in this region.

Recognizing that different trajectories can have dramatically different adaptation rates and why also has important implications for understanding the evolution of antimicrobial resistance and designing treatment strategies for malaria and other microbial pathogens. For example, maximizing the predicted time along adaptive trajectories towards antimicrobial resistance could become another objective in determining which drug and/or which dosage to use (for example, note the large differences in T_d between different drugs and different dosages in Fig. 2). Alternatively, one can imagine a scenario where compounds that target certain key alleles on known trajectories toward resistance are screened[29]. These strategies, while promising at face value, should be engaged with caution, as our results suggest that blocking specific trajectories may inadvertently open up other, potentially much faster trajectories towards resistance. To demonstrate how this might happen, we ran a simulation using the empirical fitness landscape of *P. falciparum* exposed to 10 μM pyrimethamine, but with the growth rate of genotype 0010 set to zero (thus simulating a hypothetical second drug that inhibits the growth of the 0010 genotype, a key intermediate on the greedy path towards higher resistance). Starting from the wild-type 0000, we find that this results in a 26% reduction in the number of time-steps until the optimal genotype 1110 becomes dominant

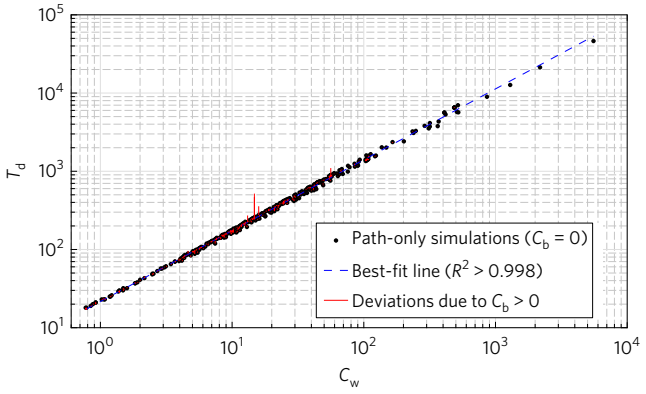

Figure 5 | Clonal interference within winning trajectories largely governs adaptation rates. Black dots represent path-only simulation times, where growth rates for all genotypes outside the path followed on the full landscape simulations have been set to zero, with the blue dashed best-fit line ($R^2 > 0.998$). Red vertical bars quantify the additional time required along the same paths on the empirical fitness landscapes, relative to on the path-only landscapes, due to C_b. C_w largely governs T_d ($n = 352$ simulations) in *P. falciparum* and *P. vivax* exposed to varying levels of pyrimethamine or cycloguanil. Only one repetition of each landscape–seed combination is shown and used to compute R^2.

(Supplementary Fig. 8). Setting the growth rate of genotype 0010 to zero in a simulation of *P. vivax* starting from the wild-type 0000 and exposed to 53.6 μM pyrimethamine has even more dramatic results, causing a 68% reduction in the number of time-steps until the optimal genotype 1110 becomes dominant (Fig. 6).

It is not clear how the timescale of our simulations relates to the evolutionary timescale of natural populations. In real malarial evolution, it is likely that the evolution along a trajectory occurs over the course of several infections under different treatment regimes (thus on different fitness landscapes), with population bottlenecks occurring at the transmission stage. Thus, the extent to which within-path and between-path competition govern adaptation rates in natural malarial populations is uncertain. Nonetheless, the dramatic differences in the speed of evolution along different trajectories, as illustrated in Fig. 6, are such that they could have a real impact on the speed of resistance evolution and should be considered when developing treatment regimes.

In this work, all of our seed populations were monomorphic. In a separate study, we recently showed how metrics of 'deception' (which quantify the degree to which sign epistasis leads a given distribution of replicators away from the global optimum) are somewhat predictive of adaptation rates in *P. falciparum* when starting from polymorphic seed populations[30]. In future work, we will investigate whether our metric for within-path competition can be combined with such measures of deceptiveness to improve convergence predictions from polymorphic populations.

The ability to predict and compare the amount of clonal interference, both within and between paths, and the insights gained by understanding the impact of such clonal interference on the resulting adaptation rates and the trajectories followed, represent new perspectives for the study of empirical fitness landscapes. However, the basic principle established here—that the speed of evolution along a given trajectory is largely governed by competition between adjacent nodes in the trajectory—has potential ramifications well beyond evolutionary biology. For example, we conjecture that the rate of adoption of evolving technologies is likely to be influenced by the amount of improvement in each new release, at least for technologies where there is fairly rapid turnover in goods, such as in software or small electronics. We believe that the study of any complex system where the fitness landscape analogy is invoked may benefit from a greater appreciation of how the greed/speed dichotomy influences innovation dynamics.

Methods

Empirical fitness landscapes. This study utilized published data from a well-characterized system of transgenic *Saccharomyces cerevesiae* carrying a combinatorially complete set of mutations (N51I, C59R, S108N and I164L) at four distinct sites in the *P. falciparum* gene for the enzyme DHFR, and the orthologous mutations in *P. vivax* (N50I, S58R, S117N, I173R). These mutations have been identified in various combinations in field isolates of malaria and are associated with drug resistance to pyrimethamine and cycloguanil. By combinatorially complete, we mean the system contains all 16 possible combinations of the absence (0) or presence (1) of each of these 4 mutations: 0000 corresponds to the wild-type ancestor. In *P. falciparum*/*P. vivax*, 1*** refers to any individual with the N51I/N50I mutation, *1** to any individual with the C59R/S58R mutation, **1* to any individual with the S108N/S117N mutation, and ***1 to any individual with the I164L/I173R mutation ('*' is a wildcard symbol that matches any value).

Using published values for drug-free growth rates and IC_{50} (the half-maximum inhibitory concentration) values for *P. falciparum* and *P. vivax* exposed to pyrimethamine[31] and for *P. falciparum* exposed to cycloguanil[32], growth rates at nine dosage levels were inferred by fitting logistic equations to these data[14,29]. The final growth rates for both species at all drug concentrations were normalized relative to the growth rate of 1011, which is the slowest-growing viable genotype of *P. falciparum* in the absence of drugs; note that the 0011 genotype has undetectable growth in the *P. falciparum* system, and is thus assigned a growth rate of zero.

This set of 464 growth rates can be considered as 29 unique four-dimensional binary empirical fitness landscapes: 19 landscapes for *P. falciparum* (one drug free and 9 levels of each of pyrimethamine and cycloguanil, as shown in the columns of Supplementary Tables 1,2) and 10 landscapes for *P. vivax* (one drug free and 9 levels of pyrimethamine, as shown in the columns of Supplementary Table 3). Each of the 16 possible genotypes for the 4 bi-allelic loci is interpreted as a location in each landscape, and their normalized growth rates as the fitness at that location. As drug levels are increased, not only do the average growth rates decrease but the landscape topography also changes[14]. For example, there are changes in the location of the globally optimal genotype (the most drug resistant in that environment, shown in bold red italic text in Supplementary Tables 1–3) as well as in the number and locations of local optima, if any (shown in blue italic text in Supplementary Tables 1–3).

Simulation model. We implemented a stochastic discrete population model that we refer to as DARPS (discrete asexually reproducing population simulator). DARPS was specifically designed to flexibly and efficiently simulate the asexual reproduction and evolution of large populations of microorganisms on complex landscapes. During each discrete time-step, the number of individuals of each genotype grows exponentially according to its particular growth rate under the current drug environment (inferred from empirical data as described above) with stochastic single-locus mutation, and then the entire population is reduced to the carrying capacity by frequency proportionate selection. The classic Wright–Fisher model[33,34] is a constant population size abstraction of the process implemented directly in DARPS, so in this sense the models are functionally equivalent. DARPS is described in more detail below.

Growth. At the start of each time-step t, the discrete number N_i^t of each genotype i in the population grows exponentially, according to its growth rate r_i under the current drug treatment regime as follows:

$$N_i^{t+1} \leftarrow \lfloor N_i^t e^{r_i} \rfloor + \text{int}(\text{rand} < (N_i^t e^{r_i} - \lfloor N_i^t e^{r_i} \rfloor)) \quad (1)$$

where $\leftarrow$ denotes variable assignment and $\lfloor \ldots \rfloor$ is the floor function. Note that the second term adds one individual of genotype i with probability $N_i^t e^{r_i} - \lfloor N_i^t e^{r_i} \rfloor$, to maintain the discrete nature of the population while still allowing genotypes with very low growth rates a small chance of increasing in number.

Mutation. For a specified probability of mutation P_m, we assume that the expected number of mutants from each genotype i is $M_i^{t+1} = N_i^{t+1} P_m$. Unless stated otherwise, the simulations reported here generated $\text{round}(M_i^{t+1})$ mutants from each i, at each t. For the comparative results shown in Supplementary Fig. 6, we also implemented this by sampling from a Poisson distribution with a mean of M_i^{t+1} to determine the number of mutants from each i. In either case, each of the selected mutants for a given genotype is stochastically assigned to mutate into one of its L neighbouring genotypes (that is, differing at exactly one locus) with equal probability, where L is the number of loci; this is the primary source of stochasticity in most of the results shown. Genotype counts N_i^{t+1} are updated accordingly after mutation.

Death by competition. Following mutation, the population size is returned to approximately K individuals using frequency proportionate selection.

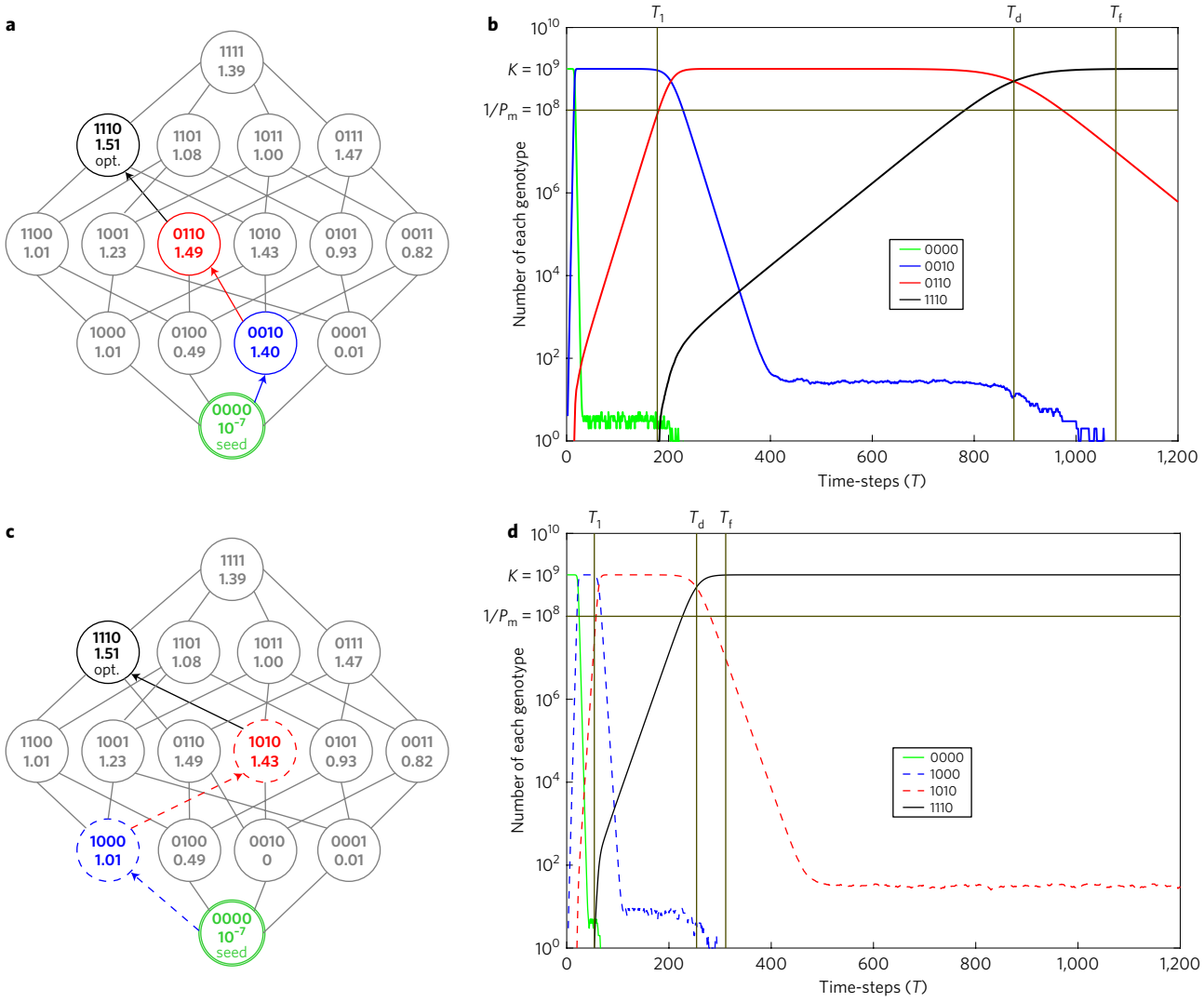

Figure 6 | Blocking the greedy path can speed up the evolution of resistance. a, Hypercube depiction of the empirical fitness landscape for *P. vivax* exposed to 53.6 μM pyrimethamine; the greedy path is followed from the wild-type 0000 to the global optimum 1110. **b**, Representative simulation on the landscape shown in **a**. **c**, The same landscape, except with the growth rate of 0010 set to 0, opens up a much faster path towards the global optimum. **d**, A representative simulation on the landscape shown in **c** converges to the global optimum much more quickly. For visual clarity, only those genotypes on the paths followed are shown.

For the results reported here we implement this using discrete stochastic renormalization for each i as follows:

$$N_i^{t+1} \leftarrow \lfloor \mathrm{freq}_i^{t+1} K \rfloor + \mathrm{int}(\mathrm{rand} < (\mathrm{freq}_i^{t+1} K - \lfloor \mathrm{freq}_i^{t+1} K \rfloor)) \quad (2)$$

where

$$\mathrm{freq}_i^{t+1} = \frac{N_i^{t+1}}{\sum_i^{2^L} N_i^{t+1}} \quad (3)$$

is the new frequency of i. The second term of equation (2) probabilistically adds one individual to maintain the discrete nature of the population while still allowing very-low-frequency genotypes a small probability of surviving. This is a very fast but relatively deterministic form of frequency proportionate selection. Alternatively, we implemented this by sampling from the current population according to a Poisson distribution with means $K\mathrm{freq}_i^{t+1}$, for each i, to better account for drift. However, for large populations such as used in this study, we confirmed that the results from these two implementations are indistinguishable ($R^2 > 0.999$, $P \approx 0$).

Termination. The population is allowed to evolve until the optimal genotype in that environment becomes dominant (more frequent than any other genotype), plus 1,000 additional time-steps to try to achieve stochastic equilibrium. If the optimal genotype does not become dominant within some prespecified number of maximum time-steps, the evolution is halted.

Computational simulations performed. For each of the 29 unique fitness landscapes, we ran simulations starting from K individuals of each of the 16 genotypes, referred to as the seed genotypes. This assumes that drug treatment begins after the population has grown to size K. We note that although malarial infections may start with a relatively small number of individuals, their populations grow exponentially in a drug-free environment until they become large enough to elicit symptoms in the host, sometimes reaching population sizes on the order of 10^{11} or 10^{12} in very ill hosts[20]. Mutation rates in the *P. falciparum* DHFR gene have been estimated to be approximately 2.5×10^{-9} per locus per replication or less[19]. As in our model P_m refers to one mutation in any of four loci per replication, except where otherwise stated, we used $P_m = 4 \times 2.5 \times 10^{-9} = 1 \times 10^{-8}$. Unless otherwise stated, the carrying capacity was set to $K = 10^9$.

All simulations were run until either the optimal genotype (the one with the highest growth rate on that landscape) became dominant or until 100,000 generations had elapsed. For each simulation we recorded: (1) at what time-step the first individual of the optimal genotype appeared (T_1); (2) at what time-step the optimal genotype became dominant (T_d); (3) at what time-step the optimal genotype became fixed (T_f, defined as >99% of the population); and (4) which trajectory (path) led to the creation of the first individual of the optimal genotype. For each identified path that was followed from the seed genotype to the optimal genotype on an empirical fitness landscape we then ran new simulations where the growth rates of all of the genotypes in the landscape that are not on the identified path were set to zero, and again recorded T_1, T_d and T_f. We refer to these as path-only simulations, where C_b has been forced to zero.

We define the greedy path as the path resulting from a 'best-first' search strategy, starting from the seed genotype. That is, if the current genotype on the path is on a fitness peak then the greedy path is terminated; otherwise the adjacent genotype with the highest fitness is added to the greedy path and becomes the current genotype, and the process is repeated. If, during 100 repetitions of a given landscape–seed experiment, the greedy path was not always the one to generate the first individual of the optimal genotype, we then ran 1,000 additional repetitions of that experiment to generate comparative statistics.

Derivation of C_w. Consider a population that comprises two genotypes that differ at exactly one locus. According to replicator dynamics[35], the rate of change of the subpopulation for genotype 2 due to competition is given by

$$\frac{dN_2}{dt} = N_2(r_2 - (r_1 p_1 + r_2 p_2)) \tag{4}$$

where r_i denotes growth rate and p_i denotes the population frequency of i. Thus for a population of size $K = N_1 + N_2$ we can write equation (4) as

$$\frac{dN_2}{dt} = N_2\left(r_2 - \left(r_1\frac{K-N_2}{K} + r_2\frac{N_2}{K}\right)\right) \tag{5}$$

If we let $y = N_2/K$, then we can write the above as

$$\frac{dy}{dt} = (r_2 - r_1)y(1-y) \tag{6}$$

This is an ordinary differential equation for the logistic function, having general solution

$$y(t) = \frac{y_0}{(1-y_0)e^{-(r_2-r_1)t} + y_0} \tag{7}$$

If we consider a population that is initially dominated by genotype 1 and then the first mutation to genotype 2 occurs at time $t = 0$, then $y(0) = 1/K$. Thus

$$y(t) = \frac{1/K}{(1-1/K)e^{-(r_2-r_1)t} + 1/K} \tag{8}$$

Assuming $r_2 > r_1$, we find the fixation time T from first introduction of genotype 2 until it reaches some prespecified fraction of the carrying capacity ρK, where ρ is arbitrarily close to 1, as

$$\rho = \frac{1/K}{(1-1/K)e^{-(r_2-r_1)T} + 1/K} \tag{9}$$

$$(1-1/K)e^{-(r_2-r_1)T} + 1/K = 1/(\rho K) \tag{10}$$

$$e^{-(r_2-r_1)T} = \frac{1/(\rho K) - 1/K}{1-1/K} \tag{11}$$

$$-(r_2-r_1)T = \ln\left(\frac{1/(\rho K) - 1/K}{1-1/K}\right) \tag{12}$$

$$T = \frac{\ln\left(\frac{1-1/K}{1/(\rho K)-1/K}\right)}{(r_2-r_1)} \tag{13}$$

Thus, we see that

$$T \propto \frac{1}{r_2 - r_1} \tag{14}$$

To generalize the effects of competition over an accessible evolutionary trajectory $(1 \to 2 \to \ldots \to m)$ of m adjacent genotypes of increasing fitness, we start with the assumption that exactly one mutant individual of genotype $i+1$ is created, and that this occurs immediately after the predecessor genotype i has reached population size ρK. Under this assumption, the time T_m to move from a population dominated by genotype 1 to one dominated by genotype m would be proportional to the summation

$$T_m \propto \sum_{i=1}^{m-1} \frac{1}{r_{i+1} - r_i} \tag{15}$$

We define this summation as the within-path clonal competition C_w along the trajectory

$$C_w = \sum_{i=1}^{m-1} \frac{1}{r_{i+1} - r_i} \tag{16}$$

The linear dependence of adaptation time T_m on C_w that is shown in equation (15) was based on the assumption that there is exactly one mutational event from genotype i to $i+1$, which occurs directly after genotype i has reached ρK individuals. Allowing for more realistic mutational events, we expect that there will be a nonlinear dependence of T_m on C_w given by

$$T_m = \alpha(C_w)^\beta \tag{17}$$

where α and β are positive constants that are functions of specific parameters of the system, including K, P_m and possibly some absolute growth rate multiplier, if the r_i values are normalized growth rates, as they are in this study (see Supplementary Table 5 for a sensitivity study). For realistic mutation rates[19] and parasite loads[20] for malaria, ongoing mutational events will speed up the evolutionary process for two reasons. First, a mutant of genotype $i+1$ may appear before i reaches ρK individuals. Second, additional ongoing mutational events will continue to increase the number of representatives of genotype $i+1$ for selection to act upon. Under these conditions, T_m will have a sublinear dependence on C_w (that is $\beta < 1$), and as the number of mutations increases, β will decrease (as shown in Supplementary Table 5). Conversely, if the number of mutations is so small that there is a delay between fixation of genotype i and the expected appearance of a mutant of genotype $i+1$, adaptation will be slowed and the relationship between T_m and C_w will be superlinear ($\beta > 1$). In practice, for observations based on path-only simulations in a given system, one can determine α and β by linear regression of $\ln(T_m)$ versus $\ln(C_w)$, where T_m is approximated by the time it takes to reach some arbitrary stopping criterion, as we have done to get the best-fit lines in Fig. 5, Supplementary Fig. 4, and Supplementary Table 5, which exhibit remarkably high coefficients of determination. However, even without knowing the specific coefficients α and β for a given system, C_w provides an easily computed and convenient metric for predicting how within-path clonal interference will affect the relative speeds of adaption along alternative accessible evolutionary trajectories, on the same or different landscapes, without the need for simulation.

Statistical analyses. Simulation times for repetitions of the same landscape–seed experiment were not always normally distributed due to lucky stochastic events that tended to create a slight negative skew. Thus, when comparing simulation times between two different experiments, we used a two-sided Wilcoxon rank sum test to compare the medians. All coefficients of determination were computed from simulation times of only one repetition of each landscape–seed experiment, to preclude inflating the R^2 values due to non-independence of repetitions of the same experiment.

Code availability. Open-source Matlab code for DARPS is available online at http://www.cs.uvm.edu/~meppstei/DARPS.

Data availability. No primary data were generated in this study. All of the published data used for the empirical fitness landscapes are available in Supplementary Tables 1–3.

Received 15 April 2016; accepted 2 September 2016; published 21 November 2016

References

1. Wright, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In *Proc. 4th Int. Congress Genetics* (ed. Jones, D. F.) Vol. 1, 356–366 (The Genetics Society of America, 1932).
2. Kauffman, S., Lobo, J. & Macready, W. G. Optimal search on a technology landscape. *J. Econ. Behav. Organ.* **43,** 141–166 (2000).
3. Holland, J. H. *Adaptation in Natural and Artificial Systems: an Introductory Analysis with Applications to Biology, Control, and Artificial intelligence* (Univ. Michigan Press, 1975).
4. Marion, R. *The Edge of Organization: Chaos and Complexity Theories of Formal Social Systems* (Sage Publications, 1999).
5. Weinreich, D. M., Delaney, N. F., DePristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312,** 111–114 (2006).
6. Lozovsky, E. R. *et al.* Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proc. Natl Acad. Sci. USA* **106,** 12025–12030 (2009).
7. de Visser, J. A. G. & Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* **15,** 480–490 (2014).
8. Poelwijk, F. J., Kiviet, D. J., Weinreich, D. M. & Tans, S. J. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445,** 383–386 (2007).
9. Palmer, A. C. & Kishony, R. Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. *Nat. Rev. Genet.* **14,** 243–248 (2013).
10. Carneiro, M. & Hartl, D. L. Adaptive landscapes and protein evolution. *Proc. Natl Acad. Sci. USA* **107,** 1747–1751 (2010).

11. Weinreich, D. M., Lan, Y., Wylie, C. S. & Heckendorn, R. B. Should evolutionary geneticists worry about higher-order epistasis? *Curr. Opin. Genet. Dev.* **3,** 700–707 (2013).
12. Tan, L., Serene, S., Chao, H. X. & Gore, J. Hidden randomness between fitness landscapes limits reverse evolution. *Phys. Rev. Lett.* **106,** 198102 (2011).
13. Jiang, P.-P., Corbett-Detig, R. B., Hartl, D. L. & Lozovsky, E. R. Accessible mutational trajectories for the evolution of pyrimethamine resistance in the malaria parasite plasmodium vivax. *J. Mol. Evol.* **77,** 81–91 (2013).
14. Ogbunugafor, C. B., Wylie, C. S., Diakite, I., Weinreich, D. M. & Hartl, D. L. Adaptive landscape by environment interactions dictate evolutionary dynamics in models of drug resistance. *PLoS Comput. Biol.* **12,** e1004710 (2016).
15. Toprak, E. *et al.* Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat. Genet.* **44,** 101–105 (2012).
16. Palmer, A. C. *et al.* Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. *Nat. Commun.* **6,** 7385 (2015).
17. De Visser, M. *et al.* Diminishing returns from mutation supply rate in asexual populations. *Science* **283,** 404–406 (1999).
18. Miralles, R., Gerrish, P. J., Moya, A. & Elena, S. F. Clonal interference and the evolution of RNA viruses. *Science* **285,** 1745–1747 (1999).
19. Paget-McNicol, S. & Saul, A. Mutation rates in the dihydrofolate reductase gene of plasmodium falciparum. *Parasitology* **122,** 497–505 (2001).
20. Dondorp, A. M. *et al.* Estimation of the total parasite biomass in acute falciparum malaria from plasma pfhrp2. *PLoS Med.* **2,** e204 (2005).
21. Gerrish, P. J. & Lenski, R. E. The fate of competing beneficial mutations in an asexual population. *Genetica* **102,** 127–144 (1998).
22. Desai, M. M., Fisher, D. S. & Murray, A. W. The speed of evolution and maintenance of variation in asexual populations. *Curr. Biol.* **17,** 385–394 (2007).
23. Jain, K., Krug, J. & Park, S.-C. Evolutionary advantage of small populations on complex fitness landscapes. *Evolution* **65,** 1945–1955 (2011).
24. Ochs, I. E. & Desai, M. M. The competition between simple and complex evolutionary trajectories in asexual populations. *BMC Evol. Biol.* **15,** 1 (2015).
25. Rozen, D. E., Habets, M. G., Handel, A. & de Visser, J. A. G. Heterogeneous adaptive trajectories of small populations on complex fitness landscapes. *PLoS ONE* **3,** e1715 (2008).
26. Kryazhimskiy, S., Rice, D. P. & Desai, M. M. Population subdivision and adaptation in asexual populations of saccharomyces cerevisiae. *Evolution* **66,** 1931–1941 (2012).
27. Nahum, J. R. *et al.* A tortoise–hare pattern seen in adapting structured and unstructured populations suggests a rugged fitness landscape in bacteria. *Proc. Natl Acad. Sci. USA* **112,** 7530–7535 (2015).
28. Nwakanma, D. C. *et al.* Changes in malaria parasite drug resistance in an endemic population over a 25-year period with resulting genomic evidence of selection. *J. Infect. Dis.* **209,** 1126–1135 (2013).
29. Ogbunugafor, C. B. & Hartl, D. A pivot mutation impedes reverse evolution across an adaptive landscape for drug resistance in *Plasmodium vivax*. *Malaria J.* **15,** 1 (2016).
30. Eppstein, M. J. & Ogbunugafor, C. B. Quantifying deception: A case study in the evolution of antimicrobial resistance. In *Proc. 2016 Genet. Evol. Comput. Conf.* 101–108 (ACM, 2016).
31. Brown, K. M. *et al.* Compensatory mutations restore fitness during the evolution of dihydrofolate reductase. *Molec. Biol. Evol.* **27,** 2682–2690 (2010).
32. Costanzo, M. S., Brown, K. M. & Hartl, D. L. Fitness trade-offs in the evolution of dihydrofolate reductase and drug resistance in *Plasmodium falciparum*. *PLoS ONE* **6,** e19636 (2011).
33. Fisher, R. A. *The Genetical Theory of Natural Selection* (Oxford Univ. Press, 1930).
34. Wright, S. Evolution in mendelian populations. *Genetics* **16,** 97–159 (1931).
35. Taylor, P. D. & Jonker, L. B. Evolutionary stable strategies and game dynamics. *Math. Biosci.* **40,** 145–156 (1978).

Acknowledgements

The authors thank J. Bagrow, C. Goodnight, S. Scarpino, D. Weinrich and J. Weitz for helpful discussions, and S. Heinrich and J. Payne for valuable comments on the manuscript. C.B.O. was supported by the Ford Foundation Postdoctoral Fellowship and the George Washington Henderson Fellowship Program at the University of Vermont.

Author contributions

C.B.O. and M.J.E. conceived the experiments, M.J.E. wrote the code, performed and analysed the experiments, and created the figures and tables. C.B.O. and M.J.E. interpreted the results and wrote the paper.

Additional information

Supplementary information is available for this paper.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.J.E.

How to cite this article: Ogbunugafor, C. B. & Eppstein, M. J. Competition along trajectories governs adaptation rates towards antimicrobial resistance. *Nat. Ecol. Evol.* **1,** 0007 (2016).

Competing interests

The authors declare no competing financial interests.