Frequency-Dependent Ecological Interactions Increase the Prevalence, and Shape the Distribution, of Preexisting Drug Resistance

Jeff Maltas [1,2,*] Dagim Shiferaw Tadele [1,3] Arda Durmaz [1] Christopher D. McFarland [2,5]
Michael Hinczewski [4] and Jacob G. Scott [1,2,4,5,†]

[1]*Cleveland Clinic, Translational Hematology Oncology Research, Cleveland, Ohio 44195, USA*
[2]*Case Western Reserve University, School of Medicine, Cleveland, Ohio 44106, USA*
[3]*Oslo University Hospital, Ullevål, Department of Medical Genetics, Oslo 0424, Norway*
[4]*Case Western Reserve University, Department of Physics, Cleveland, Ohio 44106, USA*
[5]*Case Comprehensive Cancer Center, Cleveland, Ohio 44106, USA*

The evolution of resistance remains one of the primary challenges for modern medicine, from infectious diseases to cancers. Many of these resistance-conferring mutations often carry a substantial fitness cost in the absence of treatment. As a result, we would expect these mutants to undergo purifying selection and be rapidly driven to extinction. Nevertheless, preexisting resistance is frequently observed from drug-resistant malaria to targeted cancer therapies in non-small-cell lung cancer (NSCLC) and melanoma. Solutions to this apparent paradox have taken several forms, from spatial rescue to simple mutation supply arguments. Recently, in an evolved resistant NSCLC cell line, we found that frequency-dependent ecological interactions between ancestor and resistant mutant ameliorate the cost of resistance in the absence of treatment. Here, we hypothesize that frequency-dependent ecological interactions in general play a major role in the prevalence of preexisting resistance. We combine numerical simulations with robust analytical approximations to provide a rigorous mathematical framework for studying the effects of frequency-dependent ecological interactions on the evolutionary dynamics of preexisting resistance. First, we find that ecological interactions significantly expand the parameter regime under which we expect to observe preexisting resistance. Next, even when positive ecological interactions between mutants and ancestors are rare, these resistant clones provide the primary mode of evolved resistance because even weak positive interaction leads to significantly longer extinction times. We then find that even in the case where mutation supply alone is sufficient to predict preexisting resistance, frequency-dependent ecological forces still contribute a strong evolutionary pressure that selects for increasingly positive ecological effects (negative frequency-dependent selection). Finally, we genetically engineer several of the most common clinically observed resistance mechanisms to targeted therapies in NSCLC, a treatment notorious for preexisting resistance. We find that each engineered mutant displays a positive ecological interaction with their ancestor. As a whole, these results suggest that frequency-dependent ecological effects can play a crucial role in shaping the evolutionary dynamics of preexisting resistance.

I. INTRODUCTION

The rapid, and often inevitable, evolution of therapy resistance is the primary threat to modern medicine's successful treatment of cancer and infectious disease (e.g., bacterial, viral, fungal, and parasitic infections) [1–5]. The story of resistance and treatment failure is strikingly similar across biological kingdoms. A patient is diagnosed and undergoes an initially successful treatment, only for a small resistant subclone of the original disease to relapse, resulting in treatment failure. For decades, the response to this paradigm has been the development of novel, more efficient drugs, targeting orthogonal pathways in hopes of winning the evolutionary arms race. While this response has undeniably resulted in major success stories when considering individual cancers or infections, the overall outlook for drug-resistant disease remains grim [6–9].

As a result, growing efforts have been made to study these diseases in an *evolutionary* context, whereby scientists seek to understand the ecological and evolutionary forces that seem inevitably to result in the untreatable disease state. Understanding these evolutionary forces that lead to resistance should allow scientists and physicians to not only design more effective drugs but, perhaps more crucially, design more effect *treatments*. For example, recent work has focused on improving and prolonging the efficacy of our already established drugs via optimal dose scheduling [10–12],

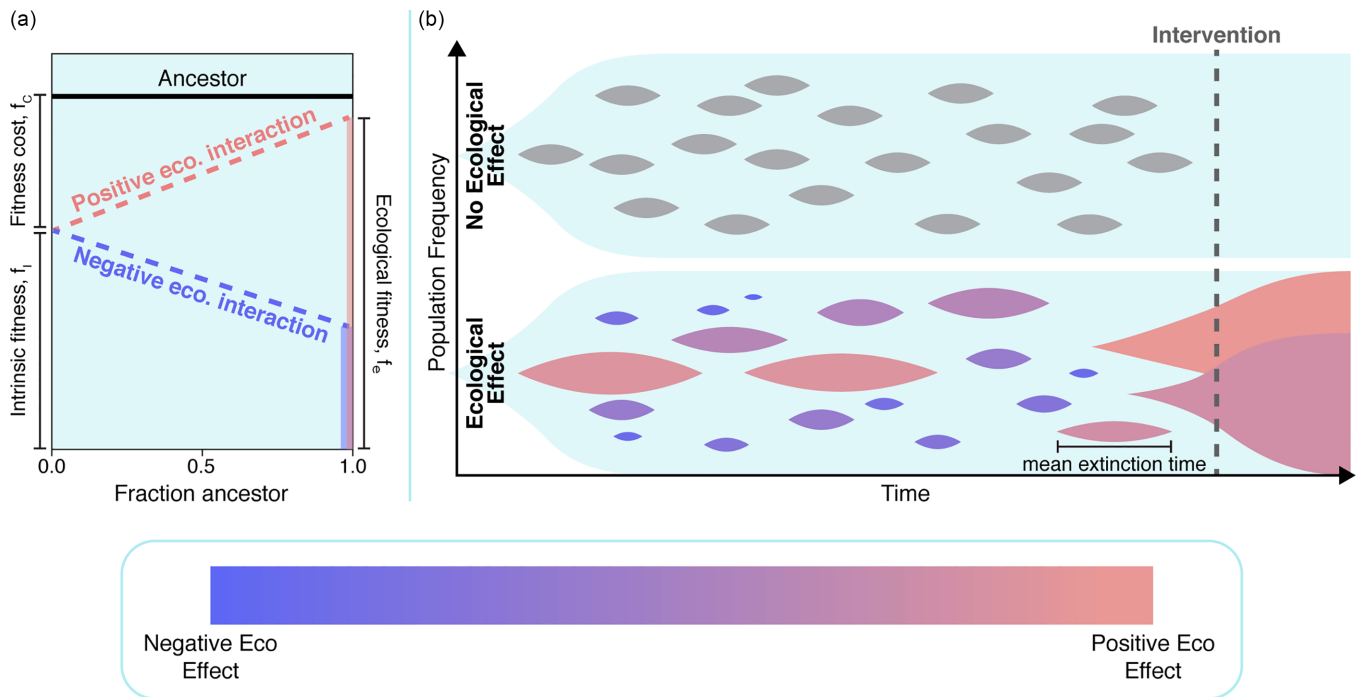*jeff.maltas@gmail.com
†scottj10@ccf.org

FIG. 1. Illustrated abstraction demonstrating how frequency-dependent ecological interactions might increase the likelihood of preexisting resistance. (a) Visualization of a typical frequency-dependent growth experiment. The ancestor (black line) is assumed to grow at a constant rate. Two hypothetical resistant mutants are depicted. Both mutants share the same intrinsic fitness and fitness cost, however the positive ecological mutant (red, growth increases as the fraction of ancestor cells increases) has a significantly higher ecological fitness ($f_e \approx 1$) than the negative ecological mutant (blue, growth decreases as the fraction of ancestor cells increases). (b) Top: Visualization of an evolving population with no ecological interactions. All mutants are assumed to have a noninsignificant fitness cost f_c, and as a result go extinct. Bottom: The same evolving population, assuming ecological interactions are present. Note that an identical number of mutants emerge, however semirare mutants with positive ecological interactions demonstrate an increased time to extinction. As a result, when a drug intervention is administered, preexistence is much more likely to be present.

drug combinations [13–17], understanding spatial dynamics [18–20], understanding ecological interactions between competing subclones [21–24], and exploiting collateral sensitivity [25–29].

In a similar spirit, this work seeks to understand the evolutionary fates of potential resistance-conferring mutations that emerge *before* treatment has occurred. The fraction of these mutants that survive to see treatment are often the primary cause of treatment failure, referred to as "preexisting resistance" [30–33]. While these resistant populations provide a large fitness advantage once treatment begins, they often carry a significant fitness disadvantage, or fitness cost (f_c), in the absence of treatment [34–38]. Nevertheless, resistance-conferring mutants often persist until treatment, at which time their treatment-sensitive ancestors are preferentially killed, resulting in the competitive release and relapse of the resistant population and inevitable treatment failure [39,40]. Understanding how these resistant clones—with a fitness disadvantage—persist in the disease population prior to treatment may allow us to prevent resistance from emerging.

This interest is derived from recent work where we measured the frequency-dependent ecological interaction between an evolved epidermal growth factor (EGFR) tyrosine kinase inhibitor (TKI) resistant non-small-cell lung cancer (NSCLC) population and its TKI-sensitive ancestor [40]. The focus of that work was on the ecological interaction under TKI treatment, and the inevitable competitive release. Strikingly, we

observed an interaction between the resistant mutant and its ancestor in the absence of any treatment. The resistant population was observed to grow about 20% slower than the ancestor when cultured separately; however, when the resistant population was cocultured with a majority ancestor population, that difference in fitness nearly vanished. This observation, referred to as negative frequency-dependent selection (negative because the fitness for the mutant decreases as the mutant frequency *increases*), is a long-studied phenomenon [41–43], and has been described as the most "intuitively obvious explanation for polymorphisms in nature" [44]. Despite its long history and potential for potent evolutionary effects, frequency-dependent selection remains understudied in the context of drug resistance. This is especially surprising, because a resistant population typically first emerges as a single individual in a predominantly ancestor population, and as a result frequency-dependent ecological interactions have a profound potential to effect the dynamics of a resistant clone (Fig. 1).

In this work, we seek to develop a rigorous theory of pretreatment evolution that incorporates frequency-dependent ecological interactions between the emerging resistant subclones and the ancestor from which they evolve. Using both a generalized Moran process and Wright-Fisher simulations, we show that mutants with the same intrinsic fitness (monoculture fitness) can have mean extinction times that vary by several orders of magnitude as a function of their ecological

fitness (fitness when cocultured in a predominantly ancestor environment). Next, we calculate the expected number of resistance-conferring mutants in the population as a function of the cost of resistance, as well as the population size and rate at which resistance-conferring mutations occur. When comparing the result of this calculation both when we assume ecological interactions exist and when they are forbidden, we identify a wide parameter space where preexistence is only likely to occur if ecological interactions are assumed. We then investigate the "many mutant regime" where preexistence is likely even without ecological interactions, and demonstrate that these ecological interactions play a prominent role in shaping the distribution of mutants, dramatically increasing the prevalence of mutants with high ecological fitness. Importantly, we show that these ecological effects drive the evolutionary outcomes even when mutants with high ecological fitness are rare. Surprisingly, despite the complexity of the model, we obtain analytical approximations for extinction rates, expected number of resistance-conferring mutants, and the distribution of observed mutants over the full range of ecological fitness. These analytical approximates both support our numerical simulations and allow us to extend our results to population sizes too large to simulate.

Finally, we test our theory experimentally by engineering several of the most common clinically observed mutations to TKI therapy in EGFR-driven NSCLC and compete these mutants against the TKI-sensitive ancestor. In all cases we observed an ecological interaction that resulted in mutant ecological fitnesses larger than their intrinsic fitness. Taken together, these theoretical and experimental results argue that frequency-dependent ecological interactions between resistant mutants and their ancestor confer a primary mode by which resistance emerges in modern cancer therapeutics, and potentially all evolutionary diseases.

II. RESULTS

A. Ecologically dependent extinction time distributions with a generalized Moran process

We begin by considering a one-step birth-death process [45–47] with states $s \in \{0, 1, \ldots, N\}$, where N is the total population size, s is the mutant population, and $N - s$ is the ancestor population. We do not consider mutation, and as a result the states $s = 0$ (extinction) and $s = N$ (fixation) are absorbing. To account for ecological interactions, the mutant's growth rate is defined to be a function of $\frac{N-s}{N}$, or the fraction of the population that is of the ancestral type, and assumed to be linear. In addition, for simplicity, we define the ancestor's growth rate to be constant and, without loss of generality, normalized to 1 (see Materials and Methods section for full model details). First, we are interested in how the distribution of extinction times differs between recently emerged mutants with identical fitness costs but distinct ecological interactions [Fig. 2(a), left]. In particular, we assume one (neutral) mutant has no ecological interaction with the ancestor, and thus $1 - f_c = f_i = f_e$ [Fig. 2(a), blue], while the comparative (positive) mutant has an interaction that ameliorates the fitness cost of the mutant at extremely large ancestor fractions, $f_e = 1$ [Fig. 2(a), red]. Here, f_c is the fitness cost or

difference between the ancestor's growth rate and the mutant's monoculture growth rate, f_i is the mutant's intrinsic fitness or monoculture growth rate, and f_e is the mutant's ecological fitness or a mutant's fitness in an otherwise purely ancestor coculture environment [definitions depicted visually in Fig. 2(a), left]. In population genetics literature this positive interaction is known as negative frequency-dependent selection and fitness is typically plotted as a function of the mutant's frequency rather than the ancestor. Throughout this work we have chosen to plot fitness relative to the ancestor's frequency in order to more intuitively connect a positive ecological interaction with a positive slope in frequency-dependent growth plots. In the case of a mutant with a positive ecological interaction, we see that the extinction time distribution is heavily right-skewed in comparison to a neutral ecological effect. As a result, if these two mutants were equally likely to emerge in a population, we would expect to observe a mutant with a positive ecological interaction significantly more often than an equivalent mutant with a neutral ecological interaction. However, ecological interactions are not always positive. Repeating this process in comparing a neutral mutant with a mutant that has a negative ecological interaction with the ancestor reveals a distinct shift to shorter extinction times (Fig. S1).

B. Extinction times depend on ecological interactions in a Wright-Fisher model

While formulating our system as a generalized Moran process allows for convenient closed-form solutions to quantities of interest such as extinction time distributions, this representation becomes computationally expensive as the population size approaches increasingly realistic values. In addition, we have completely ignored mutation, as well as more realistic conditions where many mutants are competing within an evolving population. As such, we switch to a Wright-Fisher formulation of our system [48–50]. In the Wright-Fisher model, populations are still constant in population size N, however each individual of the population is replaced every generation with offspring inheriting the parent's genotype with probability proportional to the parent's fitness. (Note that in certain cases systems with varying populations can be approximately mapped onto a Wright-Fisher model that describes similar evolutionary dynamics, in which case N is interpreted as an effective population size [51]. However, care must be taken in using this mapping in the presence of selection, where the precise form of the population change over time influences its applicability [52].) In addition, individuals acquire mutations with some probability μ and we assume mutant populations are sufficiently small that we can ignore both ecological and genetic mutant-mutant interactions. Still, several characteristics from the generalized Moran process remain. Namely, the ancestor's growth is defined to be constant and normalized to 1, and the mutant growth rate is assumed to vary linearly between f_i and f_e (as a result, a mutant's growth is fully characterized by these two fitness values along with the fraction of the population that is ancestor).

Each simulation begins with an exclusively ancestor population, and with each generation cells mutate with probability μ. Each mutant that arises has an intrinsic fitness drawn with
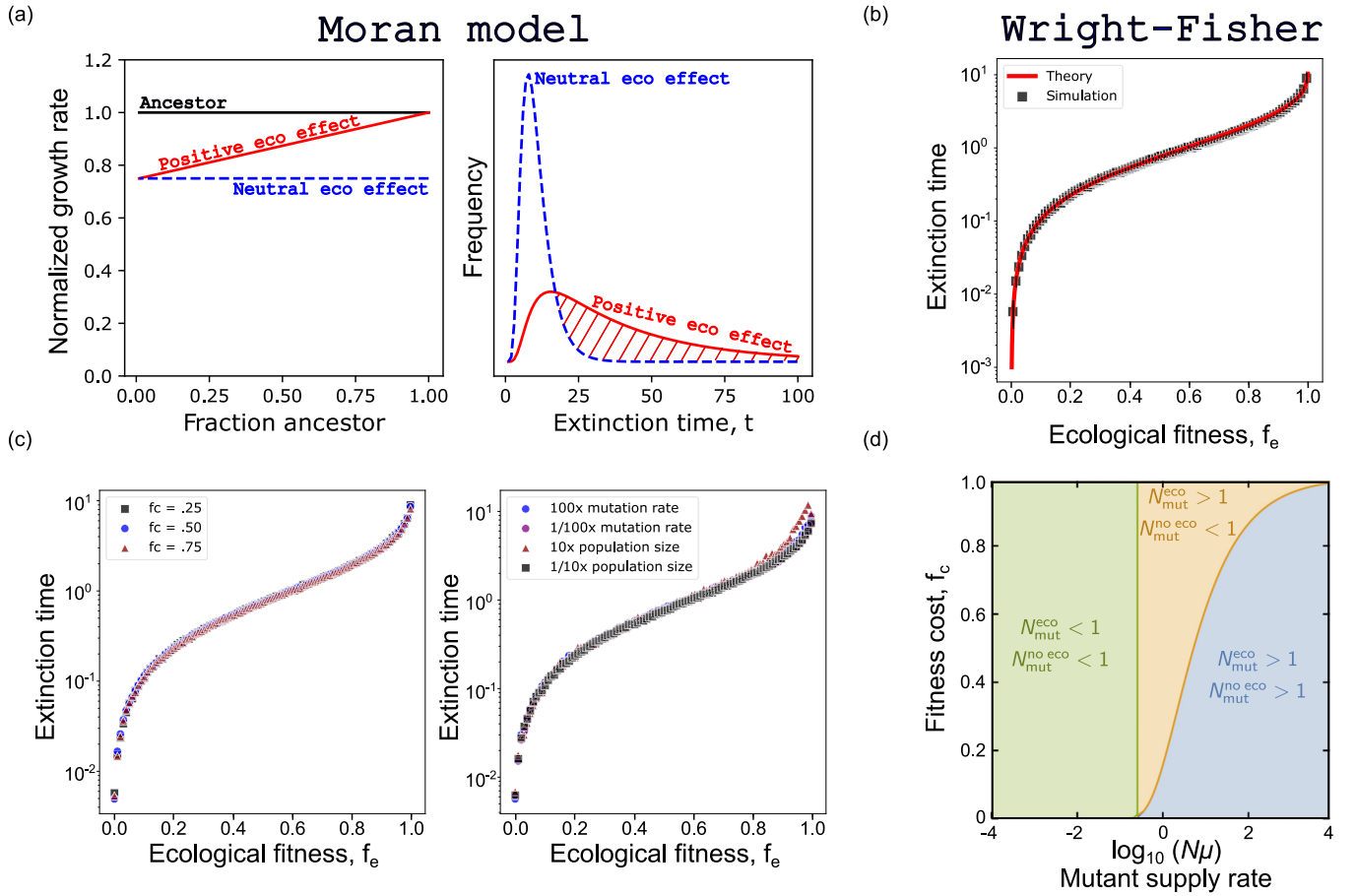
FIG. 2. Analytical approximations and simulations predict that extinction times depend on ecological interactions. (a) Closed form extinction time distributions are calculated and visualized for a generalized Moran process ($N = 100$, $f_c = 0.25$). The red distribution results from a mutant with a positive ecological interaction with the ancestor ($f_e = 1.0$), while the blue population has no ecological interaction with the ancestor ($f_e = 1 - f_c = 0.75$). (b) Wright-Fisher simulations are used to numerically calculate the mean extinction time as a function of f_e ($N = 10\,000$, $\mu = 10^{-6}$, 500 generations, f_i is drawn uniformly in $[0, 1 - f_c]$, f_i is drawn uniformly in $[0,1]$). (c) Wright-Fisher simulations are repeated for varying values of f_c, μ, and N to confirm theoretical prediction that the extinction time distribution depends only on f_e. (d) Phase diagram depicting the three regimes of preexisting resistance.

uniform probability in $[0, 1 - f_c]$ and a corresponding ecological fitness drawn with uniform probability in $[0,1]$.

Each Wright-Fisher "generation" consists of a mutation step, followed by an offspring (selection) step. For each mutant that emerges we record its intrinsic and ecological fitness values and track its evolutionary trajectory, and thus extinction time τ. A mutant that emerges but does not survive the subsequent selection step is defined to survive zero generations. Employing this model, we find that the mean extinction time varies nearly five orders of magnitude between the most positive (approximately 10 generations) and deleterious (approximately 0.001 generations) ecological interactions [Fig. 2(b)]. In order to develop a more rigorous understanding of the evolutionary dynamics, we sought an analytical approximation for the extinction time of a mutant under the same Wright-Fisher conditions. Strikingly, we find a robust approximation across the whole range of f_e:

$$\tau(f_e) \approx \frac{3\ln(1 - f_e)}{f_e^2 - 3}. \qquad (1)$$

Despite its simple form, this approximation agrees with simulation results with a typical error of 5% [Fig. 2(b), full

derivation and details found in the SI] [61]. Interestingly, the approximation is only a function of ecological fitness and not mutation rate (assuming $\mu \ll 1$), population size, or fitness cost. This finding is supported by our simulation results [Fig. 2(c)]. The dependence of τ solely on f_e in Eq. (1) is due to two factors: (i) the small total proportion of mutants in the population, which means the fitness of a mutant is approximately f_e, and (ii) at each time step (new generation) the chances of the mutants achieving a population comparable to N are vanishingly small, so the extinction probability becomes approximately independent of N.

C. Ecological interactions can increase the probability of preexisting resistance

Next we consider the model's implications for preexisting resistance. Specifically, we are interested in quantifying the expected number of mutants in an evolving population. While it might be tempting to quickly conclude that including ecological interactions will necessarily increase the probability of preexisting resistance because positive interactions will lead to longer extinction times, it is important to note that

mutants with a high intrinsic fitness are more likely to acquire a relatively deleterious ecological fitness than one that is beneficial. As such, a careful mathematical treatment is required. When the expected number of mutants in a population is low ($N_{mut} \ll 1$), potential resistance-conferring mutations are unlikely to be present at time of treatment. Contrarily, when the expected number of mutants is greater than 1, we expect treatment-threatening resistance to be present when a drug is administered. We begin adapting our analytical model to calculate the mean number of mutants (see SI for full derivation and details) [61]. To begin, we consider the case where no ecological interactions are present (f_i completely describes the growth rate of the mutant). In this case it can be shown that $N_{mut}^{no\,eco}$, the mean number of mutants ignoring ecological interactions, is

$$N_{mut}^{no\,eco} = N\mu\left(-\frac{\ln f_c}{1 - f_c} - 1\right). \quad (2)$$

Next, we seek to find an analytical approximation for N_{mut}^{eco}, the mean number of mutants assuming ecological interactions exist. In the case of a sufficiently small mutation rate, we can approximate the total mutant number as

$$N_{mut}^{eco} \approx N\mu\left(-\frac{\ln(1 - f_{max})}{f_{max}} - 1\right) \quad \text{for} \quad \mu \ll 1. \quad (3)$$

Here, f_{max} is the maximum value that f_e can take. While we can set f_{max} arbitrarily close to 1, it can never be exactly 1 for a well-defined normalization. Interestingly, for sufficiently small μ, the ratio $\frac{N_{mut}^{eco}}{N_{mut}^{no\,eco}}$ is constant with $N\mu$. While the simplicity of the approximation is appealing, unfortunately it breaks down as μ gets large. As a result, a more robust, though significantly more complex, approximation was derived (see SI for full derivation [61]):

$$N_{mut}^{eco} \approx \frac{N\mu}{f_{max}}W\left[\left(1 - f_{max} + \mu f_c^{f_c/(f_c-1)}\right)^{-1}\right]. \quad (4)$$

Here, $W(x)$ is the Lambert W function, which is the solution y of the equation $ye^y = x$. This approximation allows for efficient calculation across several decades of μ within 10% of our numerical simulations. Employing these analytical approximations, we identify three regimes of interest. The least interesting regime is the small $N\mu$ regime [Fig. 2(d), green]. Here, the effective population size is insufficient to maintain a mutant subpopulation regardless of the strength or frequency of ecological interactions. This regime corresponds to extremely rare preexisting resistance and high likelihood of treatment success.

As $N\mu$ gets larger [Fig. 2(d), yellow], we enter a regime where ecological interactions would suggest preexisting resistance is likely ($N_{mut}^{eco} > 1$), while ignoring ecological interactions would suggest preexistence is still rare ($N_{mut}^{no\,eco} < 1$). In this regime mutants have yet to become abundant, however, mutants with strong ecological interactions persist sufficiently long to threaten treatment efficacy. Representative simulation trajectories of this "rare mutant regime" are shown in Fig. 3(a). Without ecological interactions [Fig. 3(a), top panel] the mutation rate alone is insufficient to maintain a mutant subpopulation capable of threatening future treatment efficacy. However, with the in-

troduction of ecological interactions [Fig. 3(a), bottom panel, and Fig. S(2)], rare positive ecological mutants climb to significant fractions of the population, and have measurably longer extinction times that may threaten future treatments. As one might intuitively expect, the size of this regime where ecological effects drive preexistence is heavily dependent on the imposed fitness cost of resistant mutants. We find that the larger the fitness cost imposed by resistance, the larger the comparative increase provided by allowing ecological effects.

D. Ecological interactions significantly influence the distribution of mutants

Next, we consider the final regime when $N\mu$ is large [Fig. 2(d), blue]. In this regime the mutational supply is sufficiently large to self-sustain a small, resistant subpopulation, regardless of ecological interactions (that is, both $N_{mut}^{no\,eco} > 1$ and $N_{mut}^{eco} > 1$). Representative simulation trajectories of this "many mutant regime" are shown in Fig. 3(b). At first glance one might assume this regime is uninteresting. In both cases mutants are sufficiently common to threaten future treatments, albeit ecological interactions significantly increase the steady-state fraction of resistant mutants. However, the results become more interesting when we consider the shape of the resistant subpopulation distribution. In each trajectory plot, the color is proportional to the mutants' ecological fitness, with red representing an ecological fitness near 1 and blue representing an ecological fitness near 0. By inspection, it is immediately clear that the most positive ecological mutants are over-represented in the mutant population, considering they emerge with equal probability. However, we can do better and extract this relationship explicitly from our simulations [Fig. 3(c), left]. We find, similar to the impact of ecological effects on extinction times, that the frequency of a mutant spans multiple orders of magnitude as a mutant's ecological fitness varies from 0 to 1.

Extending our previous analytical work, it is straightforward to show that the stationary distribution of mutant ecological fitnesses goes as

$$P(f_e) \approx \frac{f_e\mu}{f_{max}(1 - f_e)} \quad \text{for} \quad \mu \ll 1. \quad (5)$$

The above approximation works remarkably well despite the simplicity of its form. From this equation we find that the frequency of a mutant is invariant with respect to fitness cost and population size. This is shown explicitly via numerical simulations and visualization of the joint distribution of fitness cost and ecological fitness [Fig. 2(c), right].

E. Nonuniform ecological distributions show similar qualitative results

An important context to keep in mind with the work is that up to this point we have assumed emerging mutants are assigned an ecological fitness with uniform probability in [0,1]. This assumption was not made for simplicity, but instead out of necessity. While evolutionary biologists have spent significant time both theorizing about and measuring the distribution of fitness effects (DFE), very little time has been spent quantifying either the frequency or magnitude of
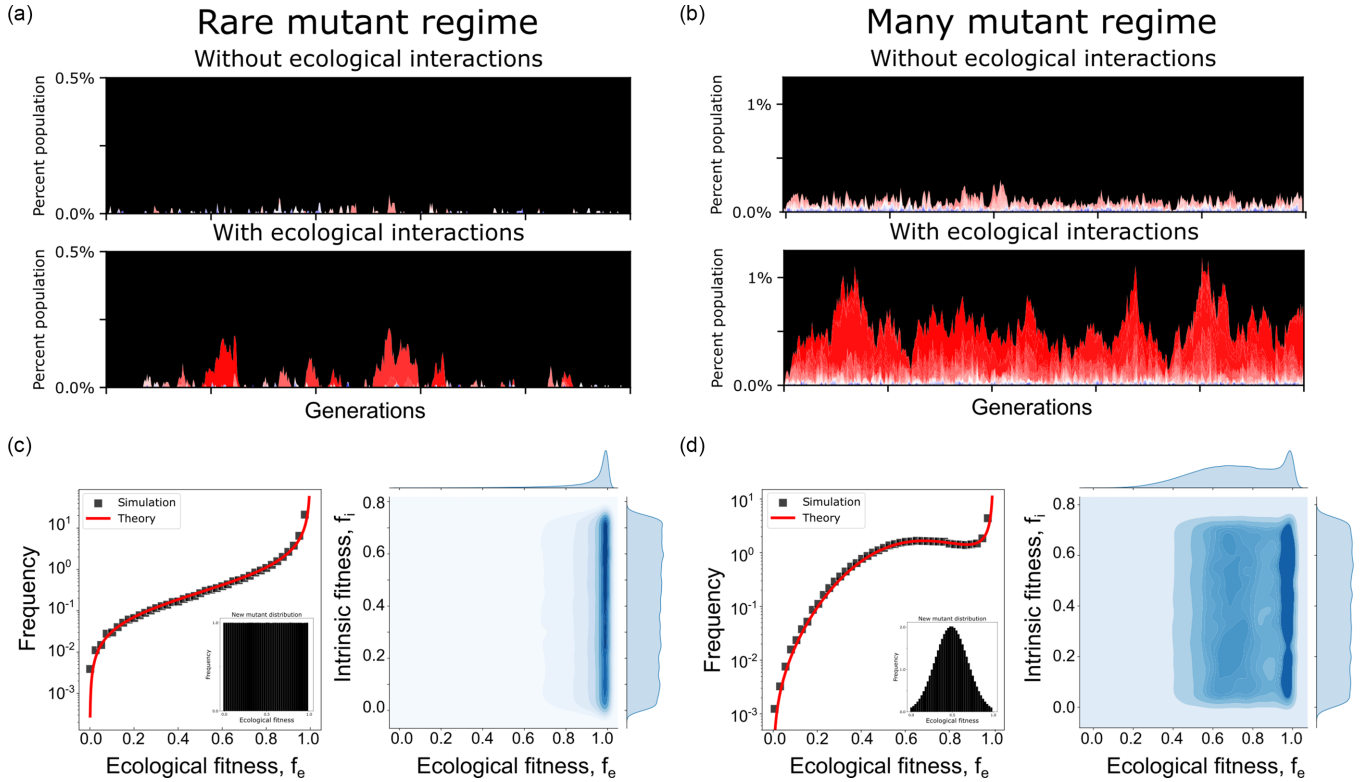
FIG. 3. Positive ecological interactions make preexistence more likely and dominate the stationary distribution of mutants. (a) Representative Wright-Fisher trajectory in the "rare mutant regime." Black corresponds to the ancestral population. Mutants exist in higher fractions and for longer periods with ecological interactions. Each mutant is colored by its ecological fitness, where red represents an f_e value near 1 and blue represents an f_e value near 0. (b) Representative trajectory in the "many mutant regime." Strong positive ecological interactions dominate the stationary distribution of mutants (visually the mutants appear red, not blue). (c) Left: Stationary distribution of mutant ecological fitnesses when the mutant-generating function is uniform across ecological fitness. Right: joint distribution density plot between intrinsic and ecological fitness. (d) Same as c, however the mutant-generating function is now Gaussian centered about $f_e = 0.5$.

ecological effects (which we propose calling the distribution of ecological effects, DEE). As a result, it is difficult to even speculate on what the null model ought to be.

Crucially, the analytical approximations derived herein can be generalized to fit any assumed, or future measured, DEE. While we assumed a uniform distribution, a Gaussian model where the most positive and negative ecological interactions are rare relative to more modest or noninteracting mutants may be more accurate. As an example, the general stationary distribution of mutant ecological fitnesses would become

$$P(f_e) \approx \frac{f_e \mu}{(1-f_e)} \rho_0(f_e) \quad \text{for} \quad \mu \ll 1. \quad (6)$$

Here, $\rho_0(f_e)$ can be any theorized or measured distribution of ecological effects. Equation (6) can be roughly interpreted as a balancing of two opposing forces to produce a stationary state: the probability $\propto f_e \mu \rho_0(f_e)$ that a new mutant with fitness f_e arises and survives is counterbalanced by the probability $\propto (1-f_e)P(f_e)$ that an existing mutant with fitness f_e disappears. As proof of principle, we numerically simulate the distribution of mutant ecological fitnesses under an assumed Gaussian DEE, and show the above analytical approximation still holds. The results are qualitatively similar to the uniform DEE and, strikingly, despite the rarity of mutants with positive ecological interactions, they still manage to dominate the predicted stationary distribution of mutants [Fig. 2(d)].

F. Sufficiently large positive ecological interactions result in a stable fixed point between mutant and ancestor

We now briefly consider the regime wherein the ecological fitness of a mutant can sample values greater than 1. Put another way, when the mutant population emerges, it may emerge into an environment where it outcompetes its ancestor. Importantly, all emerging mutants still have a nonzero fitness cost relative to the wild type and therefore as selection increases their frequency, their relative growth advantage becomes a growth disadvantage, preventing a strong selective sweep. Though our earlier analytical approximations do not apply for $f_{\max} > 1$, the numerical simulations are robust in this regime. We find that the majority of stationary distribution mutants are mutants with ecological fitnesses larger than the ancestor, or $f_e > 1$ [Fig. 4(a)]. This qualitative change in behavior above $f_e = 1$ can be explained in evolutionary game theory terms by a switch in the evolutionary game being played. When $f_e < 1$, the ancestor outcompetes the mutant population at all population frequencies. As a result, it is a question of when, not if, the mutant population will be driven to extinction. When $f_e > 1$, however, the mutant population outcompetes the ancestor at high ancestor frequencies, while the ancestor outcompetes the mutant at high mutant frequencies (as a result of the mutant fitness cost). This leads to a stable fixed point at
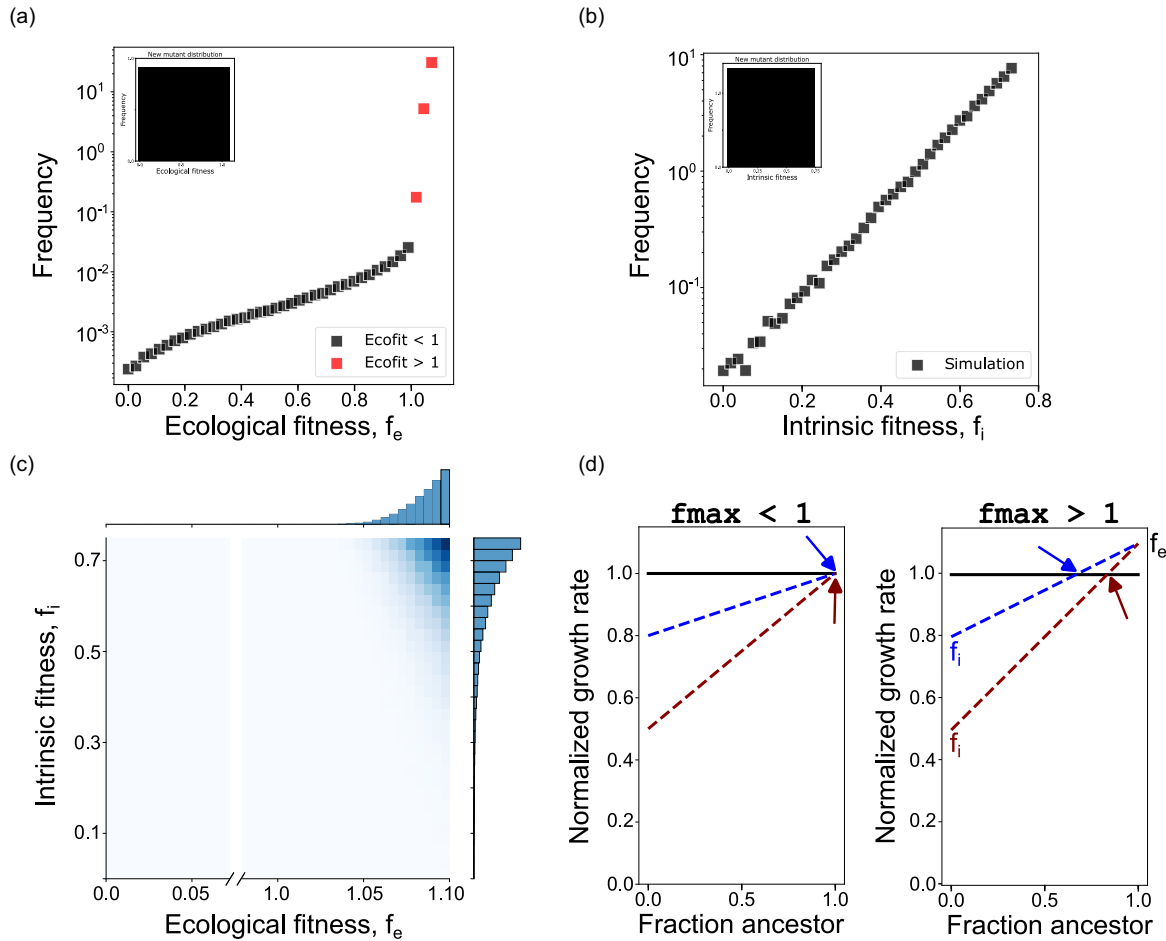
FIG. 4. Positive ecological fitnesses above 1 result in a stable fixed point between mutant and ancestor. (a) Stationary distribution of mutant ecological fitnesses when the mutant generating function has uniform probability in [0,1.10]. (b) Stationary distribution of mutant intrinsic fitnesses. (c) Joint distribution density plot between intrinsic and ecological fitness reveals the size of the fitness cost now has a significant impact on mutant survival. (d) Illustration of why two mutants with identical values of f_e can result in different extinction times. Colored arrows point to stabled fixed points between mutant and ancestor.

some ancestor frequency where the two populations have an equal growth rate. This is particularly worrying in the case of therapy-resistant mutants, because it suggests if such a mutant emerges and survives the initial stochasticity of drift, it will coexist at a sizable frequency in the population until treatment.

Next, contrary to our previous results, the fitness cost of the mutant plays an important role in determining the stationary distribution of mutants [Figs. 4(b) and 4(c)]. Here, we see that only the mutants with the largest positive ecological interactions and smallest fitness costs (intrinsic fitness = 1− fitness cost) are represented at meaningful frequencies. This result can be explained by the qualitative shift in evolutionary game for mutants where $f_e > 1$. Previously, regardless of the fitness cost, any mutant with $f_e \approx 1$ would grow at that ecological fitness, as the mutant population never became a meaningful fraction of the whole population [Fig. 4(d), left]. However, as hinted at in the numerical simulations, f_c and f_e *combine* to determine the stable fixed point between the mutant and the ancestor [Fig. 4(d), right]. As a result, even mutants with $f_e < 1$ are no longer characterized by their

ecological fitness, instead they are characterized by their fitness at the frequency determined by the stable fixed point.

G. Clinically observed lung cancer mutations confer positive ecological interactions

EGFR TKIs are the first-line treatment for patients diagnosed with advanced NSCLC. While the development of targeted TKIs has importantly extended overall survival times, these drugs are rarely curative [39] and patients often recur with TKI-resistant tumors. In addition, resistance acquisition to EGFR inhibitors has previously been linked to the selection of small preexisting mutant populations [53,54]. As a result, EGFR-mutant NSCLC is an ideal system for studying preexisting resistance and where we would expect to find positive ecological interactions between mutants and their ancestor.

To test our theory we genetically engineered (see Materials and Methods section) three of the most commonly clinically observed resistance mutations found in response to TKIs [55,56]: BRAF-V600E, KRAS-G12V, and PIK3CA-E545K. Then, using our previously described evolutionary game assay

[23,40], we measured the ecological interaction between each of these mutants and the ancestor PC9 cell line from which they emerged.

Excitingly, we found that each of the three engineered mutants had positive ecological interactions with their ancestor (Fig. 5, bottom). In addition, these positive ecological interactions are strikingly similar in both magnitude and shape to the measured ecological interaction between the evolved mutant and its ancestor (Fig. 5, top). While we already reported on the ecological interaction between the evolved mutant and its ancestor as it was the motivator of this study [40], we performed additional sequencing analysis (WXS and RNA-seq) and identified several common clinical mutations present, distinct from the engineered mutations: MET overexpression, CCND1 amplification, and KRASG12D mutation (Fig. 5, top). Taken together, the engineered and evolved mutants combine to survey approximately 70% of clinically known resistance mechanisms to TKIs in NSCLC, and in all cases we observed large positive ecological interactions ameliorating a sizable fitness cost of resistance. Next, as a control, we measured the ecological interaction between the engineered BRAF-V600E and KRAS-G12V cell lines in coculture and did not observe a positive ecological interaction (Fig. S3). This is unsurprising because a key aspect of our model is that positive ecological interactions are enriched during coevolution and thus should be more likely between ancestor and mutant, not two mutants. These experimental results are harmonious with our theoretical work and strongly support the hypothesis that frequency-dependent ecological interactions can play a critical role in the acquisition of resistance in evolutionary diseases.

III. DISCUSSION

While much work has gone into quantifying clinically problematic resistant bacteria, cancers, and viruses, we nearly always characterize these clones in monoculture—entirely outside the eco-evolutionary forces that selected for (or against) them in the first place. In this work we set out to provide the foundation for a rigorous and generalizable mathematical framework that incorporates frequency-dependent ecological interactions and can be used to study their role in preexisting resistance. This work both complements and builds off of recent studies from a wide range of disciplines ranging from theoretical population genetics and ecology to clinical trials across several biological kingdoms. We demonstrate that the presence of ecological interactions can significantly increase the probability of preexisting resistance, in addition to shaping the distribution of mutants likely to be present before treatment. We derive analytical approximations of several quantities of interest including extinction time, mean mutant population numbers, and the underlying distribution of mutants, each as a function of ecological fitness. Importantly, these results can easily be generalized to any theorized distribution of ecological effects or future experimentally measured distribution. As an important example, we show that even when we assume positive ecological interactions are rare, they still end up as a plurality of the stationary mutant frequency distribution. Finally, in a model system for preexisting resistance, we

show common clinically observed mutants harbor positive, frequency-dependent ecological interactions when cocultured with their ancestor, providing strong evidence for our theory in cancer. In addition, recent exciting work in bacteria provides additional evidence, as frequency-dependent interactions resulted in maintenance of otherwise costly antibiotic-resistant populations in *Escherichia coli* [58] and *Pseudomonas aeruginosa* [59].

It is also important to address several limitations of our work. As we mentioned earlier, the DEE has never been experimentally measured. As a result, assumptions regarding the distributional parameters have to be made in order to calculate meaningful quantities of interest. While we did our best to combat this by developing analytical models that are agnostic to this distribution, the quantitative aspect of our results are subject to the specifics of a model. Our hope is that the analytical and numerical results herein, when combined with the promising experimental work in NSCLC, motivate future measurements of the DEE across diverse model systems. Similarly, our own experimental validation is constrained to one subsystem. Our predictions are broad and should apply to many evolving populations where preexistence is evolved. Therefore it is important that future studies should should aim to test these theories not just in other cancers, but in other organisms from HIV to drug-resistant bacteria. It is possible that these principles provide the most explanatory power in cancer and bacteria, where it is common to find highly dense heterogeneous populations, in contrast to viruses, for example. Finally, while the model aims to generally capture major evolutionary forces that may underlie preexisting resistance, it is still an abstraction of a much more complex clinical scenario where the immune system, spatial dynamics, and treatment adherence, to name only a few, can play major roles.

IV. MATERIALS AND METHODS

A. Cell culture

PC-9 human adenocarcinoma cells derived from undifferentiated lung tissue were obtained from Sigma (Sigma, USA). PC-9 cells were cultured in RPMI-1640 medium supplemented with 10% heat inactivated fetal bovine serum (FBS) and 1% penicillin streptomycin solution at 37°C with humidity containing 5% CO_2. Cells were split every 4 days to maintain optimum confluency of approximately 80%–90%.

B. Engineering of mutant cell lines

To establish PC-9 cells stably expressing target genes, HEK-293T cells were cotransfected using TransIT-Lenti transfection reagent (Mirus, USA), with 500-ng psPAX2 (addgene, USA), 100-ng PMD2 (addgene, USA) and 500-ng of target genes. Viral particles were collected after 48 h and used to transduce PC-9 cells. Then, to establish ancestor PC-9 cells stably expressing nuclear localized green fluorescent protein (GFP), cells were transduced with pLVX-eGFP-Hygro (Vectorbuild, USA). In addition, to establish query cells expressing fluorescently labeled PC-9 cells with a gene of interest, cells were cotransduced with pLVX-mCherry-Hygro or pLVX-mCherry-Puro and pLVX-PIK3CA-E545K-Bsd (Vectorbuild, USA). Next, 72 h after transduction, cells were selected with
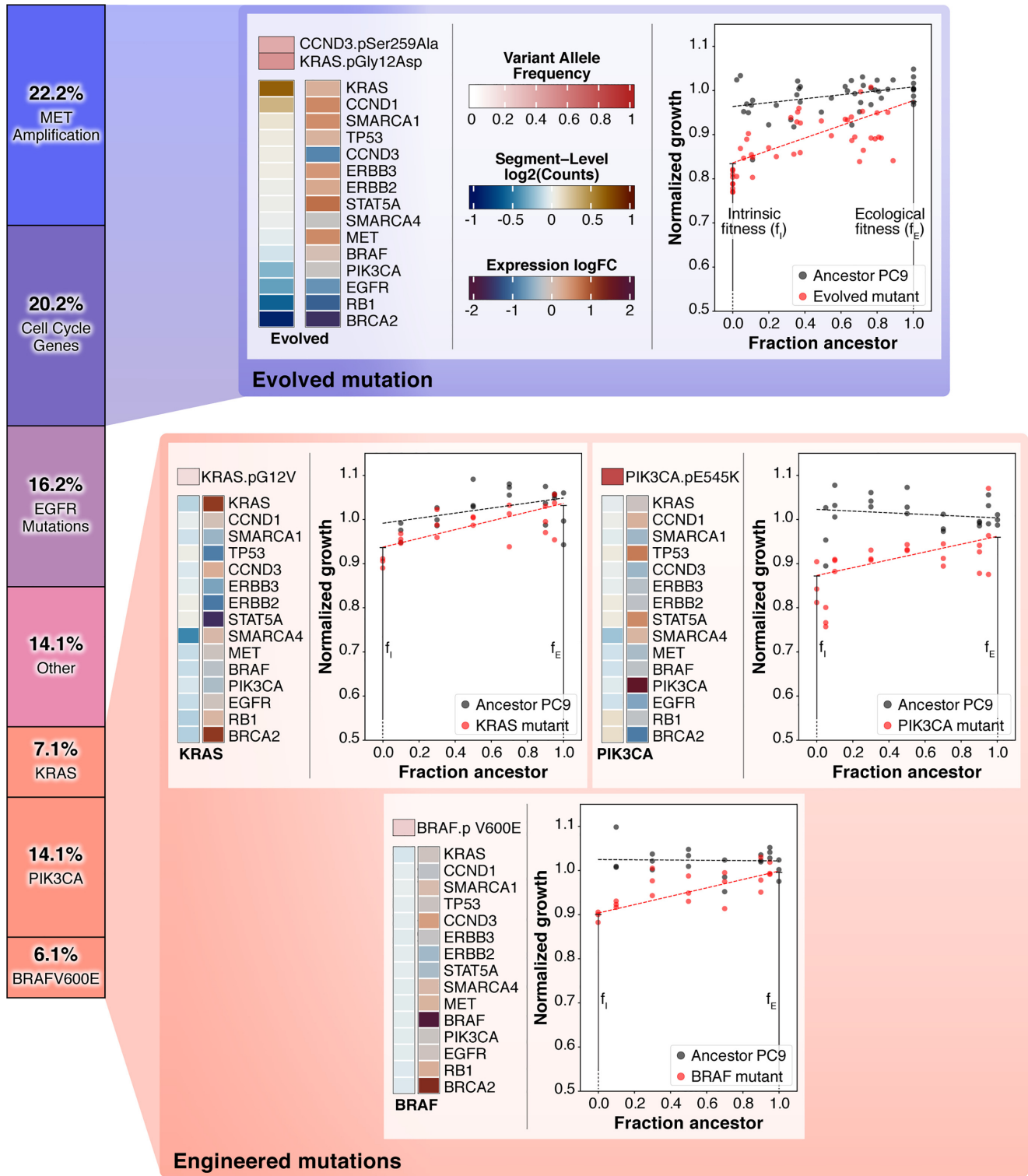
FIG. 5. Common clinically observed resistance mutations in NSCLC harbor strong positive ecological interactions with their ancestor in a model system of preexisting resistance. Stacked bar chart: Visual representation of the known resistance mechanisms to osimertinib, a third-generation TKI and the current standard of care for EGFR-positive NSCLC. Mutation frequencies and categorical definitions from Leonetti *et al.* [57]. Top: Evolved gefitinib-resistant NSCLC PC9 mutant (previously reported [40]) exhibits a positive ecological interaction with its ancestor. Fresh sequencing analysis identifies clinically observed resistant mutations including KRASG12D, MET amplification, and CCND1 amplification (cell cycle gene). Bottom: Measured positive ecological interactions between engineered resistant mutants (KRAS-G12V top left, PIK3CA-E545K top right, BRAF-V600F bottom) and their ancestor.

200 µg/mL hygromycin, 5 µg/mL puromycin, and 5 µg/mL blasticidin.

C. Drug sensitivity assay

Cells were harvested at 70%–80% confluence, stained with trypan blue (Corning, USA), and counted with a TC20 Automated Cell Counter (Bio-Rad, USA). Luminescent-based cell viability assays using CellTiter-Glo (CTG) reagent (Promega, USA) were performed in a 96-well plate (Corning, USA). A total of 3000 cells were plated in 90 µL of complete medium per well in three replicates per drug concentration with a Multidrop reagent dispenser (Thermo Fishers, USA). After 3 h of incubation, 10 µL of gefitinib, osimertinib, and erlotinib (Cayman, USA) diluted in complete RPMI-1640 medium were added to the cells. Compounds were tested in a threefold dilution in a range of 0–1.8 µM, 0–3 µM, and 0–10 µM for gefitinib, osimertinib, and erlotinib, respectively. After 72 h of incubation, 25 µL of CTG reagent was add to the cells, then incubated for 10 min at room temperature, and luminescence was measured.

D. Game assay

PC-9 mutants stably expressing nuclear localized fluorescent signal ancestor PC-9 stably expressing nuclear localized GFP were cocultured at different initial proportions of ancestor cells at a density of 1500 cells in 90 µL of fresh medium. After 3 h of incubation, 10 µL of Dimethyl sulfoxide (DMSO) diluted in complete RPMI-1640 medium (final DMSO concentration of 0.1% v/v) were added to the cells in three replicates per initial proportion. Then time-lapse microscopy images were obtained for GFP and mCherry using BioSpa automated incubator (BioTek, USA) every 4 h over the course of 96 h. Then, images were processed with the open-source software CELLPROFILER [60]. Images were background subtracted, converted to 8-bit, contrast enhanced, and thresholded, then raw cell numbers were extracted.

E. Generalized Moran model details

In this work we consider a well-known generalized Moran process, a model previously used to study frequency-dependent evolutionary dynamics [45–47]. Briefly, we consider a one-step birth-death process with states $s \in \{0, 1, \ldots, N\}$ and characterized by birth and death rates b_i and d_i with $i \in \{1, 1, \ldots, N-1\}$. As a result, this model describes a fixed population size N, with s resistant mutants and $N-s$ ancestor population. We forgo mutation rate and as a consequence the states $s = 0$ and $s = N$ are absorbing. We consider an evolutionary game with a 2×2 payoff matrix such that

$$
\begin{array}{c}
\quad \text{R} \qquad \text{A} \\
\begin{array}{c} \text{R} \\ \text{A} \end{array}
\begin{pmatrix} (1 - f_c) & f_e \\ 1 & 1 \end{pmatrix}.
\end{array}
$$

As a result, we can write the expected payoffs in a population of s mutants and $N - s$ ancestor individuals as

$$
P(R) = \frac{s-1}{N-1}(1 - f_c) + \frac{N-s}{N-1}f_e \quad P(A) = 1.
$$

F. RNA-Seq

Paired-end reads are preprocessed using FASTP to trim and quality filter the reads. Following the filtering, reads are aligned to the GRCh38 reference genome via a STAR aligner. Read quantification is done using SALMON on the extracted transcriptome locations from spliced STAR alignment. Gene-level abundance is aggregated from bootstrapped transcript abundances using the R package TXIMPORT. The ARRIBA tool is coupled with spliced alignments for fusion-transcript detection as well. Pathway level expression activities are quantified using the R package GSVA and MSIGDBR for the hallmark pathways. The R package COMPLEXHEATMAP was used to generate heat maps.

G. Whole exome sequencing

Paired-end whole-exome reads of ancestor and parental lines were preprocessed using FASTP similar to RNA-Seq. Alignment to the GATK (GATK best practices bucket) version of the GRCh38 reference is done using a BWA-MEM aligner. Following the alignment, variant calling pipeline according to the GATK workflow, including duplicate marking and variant calling via HAPLOTYPECALLER, was conducted. Variants passing filtering based on hard-filtering are further annotated using the variant effect predictor (VEP) tool. Exome alignments are further input to CNVKIT for copy-number alterations. Using a flat reference for bias correction, log2 scaled abundances are generated for ancestor and resistant strains. Copy-number segments are captured using circular binary segmentation and assigned to genes mapping to the segment.

H. Code availability

Code used in this study is made openly available on GITHUB (https://github.com/jamaltas/Preexistence).

[1] M. C. Fisher *et al.*, Tackling the emerging threat of antifungal resistance to human health, Nat. Rev. Microbiol. **20**, 557 (2022).

[2] D. E. Goldberg, R. F. Siliciano, and W. R. Jacobs, Outwitting evolution: Fighting drug-resistant TB, malaria, and HIV, Cell **148**, 1271 (2012).

[3] E. M. Darby *et al.*, Molecular mechanisms of antibiotic resistance revisited, Nat. Rev. Microbiol. **21**, 280 (2023).

[4] P. S. Pennings, HIV drug resistance: Problems and perspectives, infectious disease reports, Infect. Dis. Rep. **5**, e5 (2013).

[5] D. Menard and A. Dondorp, Antimalarial drug resistance: A threat to malaria elimination, Cold. Spring. Harb. Perspect. Med. **7**, a025619 (2017).

[6] A. Sandru, S. Voinea, E. Panaitescu, and A. Blidaru, Survival rates of patients with metastatic malignant melanoma, J. Med. Life **7**, 572 (2014).

[7] H. W. Boucher *et al.*, Bad bugs, no drugs: No ESKAPE! An update from the Infectious Diseases Society of America, Clin. Infect. Dis. **48**, 1 (2009).

[8] W. H. Organization, *Antimicrobial Resistance: Global Report on Surveillance* (World Health Organization, Geneva, 2014).

[9] M. Arnold *et al.*, Progress in cancer survival, mortality, and incidence in seven high-income countries 1995–2014 (ICBP SURVMARK-2): A population-based study, Lancet Oncol. **20**, 1493 (2019).

[10] J. Maltas *et al.*, Drug dependence in cancer is exploitable by optimally constructed treatment holidays, Nat. Ecol. Evol. **8**, 147 (2024).

[11] S. Iram *et al.*, Controlling the speed and trajectory of evolution with counterdiabatic driving, Nat. Phys. **17**, 135 (2021).

[12] R. A. Gatenby, A. S. Silva, R. J. Gillies, and B. R. Frieden, Adaptive therapy, Cancer Res. **69**, 4894 (2009).

[13] A. F. Feder, K. N. Harper, C. J. Brumme, and P. S. Pennings, Understanding patterns of HIV multi-drug resistance through models of temporal and spatial drug heterogeneity, eLife **10**, e69032 (2021).

[14] G. R. Anderson *et al.*, A landscape of therapeutic cooperativity in KRAS mutant cancers reveals principles for controlling tumor evolution, Cell. Rep. **20**, 999 (2017).

[15] M. Hegreness, N. Shoresh, D. Damian, D. Hartl, and R. Kishony, Accelerated evolution of resistance in multidrug environments, Proc. Natl. Acad. Sci. **105**, 13977 (2008).

[16] J. P. Torella, R. Chait, and R. Kishony, Optimal drug synergy in antimicrobial treatments, PLoS Comput Biol. **6**, e1000796 (2010).

[17] Z. Dean, J. Maltas, and K. B. Wood, Antibiotic interactions shape short-term evolution of resistance in *E. faecalis*, PLoS Pathog. **16**, e1008278 (2020).

[18] F. Fu, M. A. Nowak, and S. Bonhoeffer, Spatial heterogeneity in drug concentrations can facilitate the emergence of resistance to cancer therapy, PLoS Comput. Biol. **11**, e1004142 (2015).

[19] Q. Zhang *et al.*, Acceleration of emergence of bacterial antibiotic resistance in connected microenvironments, Science **333**, 1764 (2011).

[20] M. G. De Jong and K. B. Wood, Tuning spatial profiles of selection pressure to modulate the evolution of drug resistance, Phys. Rev. Lett. **120**, 238102 (2018).

[21] K. S. Korolev, J. B. Xavier, and J. Gore, Turning ecology and evolution against cancer, Nat. Rev. Cancer **14**, 371 (2014).

[22] A. R. Wargo, S. Huijben, J. C. De Roode, J. Shepherd, and A. F. Read, Competitive release and facilitation of drug-resistant parasites after therapeutic chemotherapy in a rodent malaria model, Proc. Natl. Acad. Sci. USA **104**, 19914 (2007).

[23] A. Kaznatcheev, J. Peacock, D. Basanta, A. Marusyk, and J. G. Scott, Fibroblasts and alectinib switch the evolutionary games played by non–small cell lung cancer, Nat. Ecol. Evol. **3**, 450 (2019).

[24] B. H. Good, M. J. McDonald, J. E. Barrick, R. E. Lenski, and M. M. Desai, The dynamics of molecular evolution over 60,000 generations, Nature (London) **551**, 45 (2017).

[25] J. Maltas and K. B. Wood, Pervasive and diverse collateral sensitivity profiles inform optimal strategies to limit antibiotic resistance, PLoS Biol **17**, e3000515 (2019).

[26] J. Maltas, A. Huynh, and K. B. Wood, Dynamic collateral sensitivity profiles highlight challenges and opportunities for optimizing antibiotic sequences, bioRxiv (2021).

[27] J. A. Scarborough *et al.*, Identifying states of collateral sensitivity during the evolution of therapeutic resistance in Ewing's sarcoma, iScience **23**, 101293 (2020).

[28] J. Maltas, D. M. McNally, and K. B. Wood, Evolution in alternating environments with tunable interlandscape correlations, Evolution **75**, 10 (2021).

[29] B. Zhao *et al.*, Exploiting temporal collateral sensitivity in tumor clonal evolution, Cell **165**, 234 (2016).

[30] J. J. Lin and A. T. Shaw, Resisting resistance: Targeted therapies in lung cancer, Trends Cancer **2**, 350 (2016).

[31] J. Boshuizen *et al.*, Reversal of pre-existing NGFR-driven tumor and immune therapy resistance, Nat. Commun. **11**, 3946 (2020).

[32] P. K. Jangir *et al.*, Pre-existing chromosomal polymorphisms in pathogenic E. coli potentiate the evolution of resistance to a last-resort antibiotic, eLife **11**, e78834 (2022).

[33] M. Robinson, Y. Tian, W. E. Delaney IV, and A. E. Greenstein, Preexisting drug-resistance mutations reveal unique barriers to resistance for distinct antivirals, Proc. Natl. Acad. Sci. USA **108**, 10290 (2011).

[34] G. Szakacs *et al.*, Targeting the Achilles heel of multidrug-resistant cancer by exploiting the fitness cost of resistance, Chem. Rev. **114**, 5753 (2014).

[35] M. E. S. Sampah, L. Shen, B. L. Jilek, and R. F. Siliciano, Dose-response curve slope is a missing dimension in the analysis of HIV-1 drug resistance, Proc. Natl. Acad. Sci. USA **108**, 7613 (2011).

[36] A. H. Melnyk, A. Wong, and R. Kassen, The fitness costs of antibiotic resistance mutations, Evol. Appl. **8**, 273 (2015).

[37] D. M. Weinreich, N. F. Delaney, M. A. DePristo, and D. L. Hartl, Darwinian evolution can follow only very few mutational paths to fitter proteins, Science **312**, 111 (2006).

[38] D. Kühnert *et al.*, Quantifying the fitness cost of HIV-1 drug resistance mutations through phylodynamics, PLoS Pathog. **14**, e1006895 (2018).

[39] H. Hayashi *et al.*, Overall treatment strategy for patients with metastatic NSCLC with activating EGFR mutations, Clin. Lung Cancer **23**, e69 (2022).

[40] N. Farrokhian *et al.*, Measuring competitive exclusion in non–small cell lung cancer, Sci. Adv. **8**, eabm7212 (2022).

[41] F. J. Ayala and C. A. Campbell, Frequency-dependent selection, Annu. Rev. Ecol. Syst. **5**, 115 (1974).

[42] B. C. Clarke, P. Shelton, and G. Mani, Frequency-dependent selection, metrical characters and molecular evolution, Philos. Trans. R. Soc. B **319**, 631 (1988).

[43] B. Levin, Frequency-dependent selection in bacterial populations, Philos. Trans. R. Soc. B **319**, 459 (1988).

[44] M. V. Trotter and H. G. Spencer, Frequency-dependent selection and the maintenance of genetic variation: Exploring the parameter space of the multiallelic pairwise interaction model, Genetics **176**, 1729 (2007).

[45] A. Traulsen, J. C. Claussen, and C. Hauert, Coevolutionary dynamics in large, but finite populations, Phys. Rev. E **74**, 011901 (2006).

[46] P. Ashcroft, A. Traulsen, and T. Galla, When the mean is not enough: Calculating fixation time distributions in birth-death processes, Phys. Rev. E **92**, 042154 (2015).

[47] M. A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg, Emergence of cooperation and evolutionary stability in finite populations, Nature (London) **428**, 646 (2004).

[48] L. A. Imhof and M. A. Nowak, Evolutionary game dynamics in a Wright-Fisher process, J. Math. Biol. **52**, 667 (2006).

[49] S. Wright, Evolution in Mendelian populations, Genetics **16**, 97 (1931).

[50] R. A. Fisher, XXI.—On the dominance ratio, Proc. R. Soc. Edinb. **42**, 321 (1923).

[51] P. Sjodin, I. Kaj, S. Krone, M. Lascoux, and M. Nordborg, On the meaning and existence of an effective population size, Genetics **169**, 1061 (2005).

[52] S. P. Otto and M. C. Whitlock, The probability of fixation in populations of changing size, Genetics **146**, 723 (1997).

[53] H.-E. C. Bhang *et al.*, Studying clonal dynamics in response to cancer therapy using high-complexity barcoding, Nat. Med. **21**, 440 (2015).

[54] A. N. Hata *et al.*, Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition, Nat. Med. **22**, 262 (2016).

[55] J. Chmielecki *et al.*, Candidate mechanisms of acquired resistance to first-line osimertinib in EGFR-mutated advanced non-small cell lung cancer, Nat. Commun. **14**, 1070 (2023).

[56] D. R. Camidge, W. Pao, and L. V. Sequist, Acquired resistance to TKIS in solid tumours: Learning from lung cancer, Nat. Rev. Clin. Oncol. **11**, 473 (2014).

[57] A. Leonetti *et al.*, Resistance mechanisms to osimertinib in EGFR-mutated non-small cell lung cancer, Br. J. Cancer **121**, 725 (2019).

[58] A. M. Leale and R. Kassen, The emergence, maintenance, and demise of diversity in a spatially variable antibiotic regime, Evol. Lett. **2**, 134 (2018).

[59] T. Dimitriu *et al.*, Negative frequency dependent selection on plasmid carriage and low fitness costs maintain extended spectrum β-lactamases in *Escherichia coli*, Sci. Rep. **9**, 17211 (2019).

[60] D. R. Stirling *et al.*, Cellprofiler 4: Improvements in speed, utility and usability, BMC Bioinf. **22**, 1 (2021).

[61] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PRXLife.2.023010 for complete derivations and additional figures.