

# On the evolutionary ecology of multidrug resistance in bacteria

Sonja Lehtinen, François Blanquart, Marc Lipsitch, Christophe Fraser

▶ **To cite this version:**

Sonja Lehtinen, François Blanquart, Marc Lipsitch, Christophe Fraser. On the evolutionary ecology of multidrug resistance in bacteria. PLoS Pathogens, 2019, 15 (5), pp.e1007763. 10.1371/journal.ppat.1007763 . hal-02347472

**HAL Id: hal-02347472**
**https://hal.science/hal-02347472**

Submitted on 11 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PLOS PATHOGENS

RESEARCH ARTICLE

# On the evolutionary ecology of multidrug resistance in bacteria

**Sonja Lehtinen[1]*, François Blanquart[1,2,3], Marc Lipsitch[4,5], Christophe Fraser[1], with the Maela Pneumococcal Collaboration[¶]**

**1** Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom, **2** Center for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS, INSERM, PSL Research University, Paris, France, **3** IAME, UMR 1137, INSERM, Université Paris Diderot, Paris, France, **4** Center for Communicable Disease Dynamics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, United States of America, **5** Departments of Epidemiology and Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, United States of America

¶ Membership of the Maela Pneumococcal Collaboration is listed in the Acknowledgments.
* sonja.lehtinen@bdi.ox.ac.uk

Check for updates

OPEN ACCESS

**Citation:** Lehtinen S, Blanquart F, Lipsitch M, Fraser C, with the Maela Pneumococcal Collaboration (2019) On the evolutionary ecology of multidrug resistance in bacteria. PLoS Pathog 15(5): e1007763. https://doi.org/10.1371/journal.ppat.1007763

**Editor:** Erwin Schurr, McGill University, CANADA

**Received:** October 30, 2018

**Accepted:** April 15, 2019

**Published:** May 13, 2019

**Copyright:** © 2019 Lehtinen et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its Supporting Information files.

**Funding:** SL and CF were funded by MIDAS (U01GM110721-01). FB was funded by Marie Sklodowska-Curie Actions (657768). ML was funded by National Institute Of General Medical Sciences (U54GM088558). FB and CF also received funding from the Li Ka Shing foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Resistance against different antibiotics appears on the same bacterial strains more often than expected by chance, leading to high frequencies of multidrug resistance. There are multiple explanations for this observation, but these tend to be specific to subsets of antibiotics and/or bacterial species, whereas the trend is pervasive. Here, we consider the question in terms of strain ecology: explaining why resistance to different antibiotics is often seen on the same strain requires an understanding of the competition between strains with different resistance profiles. This work builds on models originally proposed to explain another aspect of strain competition: the stable coexistence of antibiotic sensitivity and resistance observed in a number of bacterial species. We first identify a partial structural similarity in these models: either strain or host population structure stratifies the pathogen population into evolutionarily independent sub-populations and introduces variation in the fitness effect of resistance between these sub-populations, thus creating niches for sensitivity and resistance. We then generalise this unified underlying model to multidrug resistance and show that models with this structure predict high levels of association between resistance to different drugs and high multidrug resistance frequencies. We test predictions from this model in six bacterial datasets and find them to be qualitatively consistent with observed trends. The higher than expected frequencies of multidrug resistance are often interpreted as evidence that these strains are out-competing strains with lower resistance multiplicity. Our work provides an alternative explanation that is compatible with long-term stability in resistance frequencies.

## Author summary

Antibiotic resistance is a serious public health concern, yet the ecology and evolution of drug resistance are not fully understood. This impacts our ability to design effective interventions to combat resistance. From a public health point of view, multidrug resistance is particularly problematic because resistance to different antibiotics is often seen on the

**Competing interests:** ML has received consulting fees/honoraria from Merck, Pfizer, Affinivax, and Antigen Discovery, Inc and grant support not related to this paper from Pfizer and PATH Vaccine Solutions.

same bacterial strains, which leads to high frequencies of multidrug resistance and limits treatment options. This work seeks to explain this trend in terms of strain ecology and the competition between strains with different resistance profiles. Building on recent work exploring why resistant bacteria are not out-competing sensitive bacteria, we show that models originally proposed to explain this observation also predict high multidrug resistance frequencies. These models are therefore a unifying explanation for two pervasive trends in resistance dynamics. In terms of public health, the implication of our results is that new resistances are likeliest to be found on already multidrug resistant strains and that changing patterns of prescription may not be enough to combat multidrug resistance.

## Introduction

Antibiotic resistance and, in particular, multidrug resistance (MDR) are public health threats. Multidrug resistant infections are associated with poorer clinical outcomes and higher cost of treatment than other infections [1, 2] and there is concern that the emergence of pan-resistant strains (pathogens resistant to all available antibiotics) will render some infections untreatable [3].

From the point of view of finding effective treatment options, multidrug resistance is particularly problematic because resistance to different antibiotics tends to be concentrated on the same strains: positive correlations between resistance to different drugs have been found in multiple species (including *Streptococcus pneumoniae*, *Neisseria gonorrhoeae*, *Staphylococcus aureus*, *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Mycobacterium tuberculosis*) [2]. In other words, the frequency of MDR strains is higher than we would expect from the frequencies of individual resistance determinants if these were distributed randomly in the population ('MDR over-representation').

Understanding the causes of this MDR over-representation is important for limiting the impact of resistance. A number of possible explanations have been suggested (Table 1) [2], but the extent to which these processes contribute to the trend remains uncertain. Many of the proposed mechanisms are specific to subsets of antibiotics and/or species. The pattern of MDR over-representation, on the other hand, is pervasive: correlations have been observed between resistance to antibiotics acting through different mechanisms, and between chromosomal and mobile genetic element (MGE) associated resistance determinants [2]. Explanations for MDR over-representation must therefore be either sufficiently general or sufficiently diverse to account for this pervasiveness.

In this paper, we approach the problem of explaining MDR over-representation in terms of strain ecology: explaining why resistance to different antibiotics is often seen on the same strain requires an understanding of the competition between strains with different resistance profiles. For models of such competition to be credible, they must capture observed trends in resistance dynamics whilst being ecologically plausible. Developing models that fulfil these criteria has not been trivial: sensitive and resistant strains compete for the same hosts and simple models of competition therefore predict that the fitter strain will out-compete the other ('competitive exclusion') [19]. However, this is rarely observed: resistance frequencies have remained intermediate over long time periods in a number of species. For example, sustained intermediate resistance frequencies are observed in Europe for various antibiotics and numerous species, including *E. coli*, *S. aureus* and *S. pneumoniae* (European Centre for Disease Prevention and Control Surveillance Atlas, available at https://atlas.ecdc.europa.eu). Stable coexistence is also observed in surveillance data from multiple other locations (Centre for

**Table 1. Processes which may contribute to MDR over-representation.**

| Process | Notes |
|---|---|
| Shared resistance mechanisms | Particularly relevant for antibiotics of the same class (e.g. $\beta$-lactamases and some penicillin-binding protein mutations conferring resistance against multiple $\beta$-lactams), but also applicable for some drugs of different classes: there are examples of efflux pumps acting on multiple drugs in numerous species [4] and evidence for clinical relevance of efflux pumps in multidrug resistance [5]. However, MDR over-representation is also observed where efflux pumps are not thought to be a major mechanism of resistance (e.g. between $\beta$-lactams and other classes of antibiotics in *S. pneumoniae* [2]). |
| Linkage between resistance genes (when resistance is associated with particular alleles) | In general terms, linkage (i.e. being inherited together) between alleles is not a mechanism that generates association between alleles. However, linkage slows down the rate at which recombination breaks down associations between alleles [6], it may therefore play a role in temporarily maintaining associations between resistances. In this context, it is helpful to distinguish between resistance mechanisms where a particular allele of a gene confers resistance (e.g. changes to the protein targeted by the antibiotics) and those where resistance is associated with the presence of a resistance gene (e.g. enzymes that break down the drug). When resistance is associated with specific alleles, there is no *a priori* reason to expect the resistance allele for one antibiotic to be linked to the resistance, rather than sensitivity, allele of another antibiotic. |
| Linkage between resistance genes (when resistance is associated with presence of gene) | When resistance is associated with presence of a particular gene, absence of the gene from a particular MGE does not necessarily imply sensitivity to the antibiotic (the gene may be present on another element). As a consequence, spread of the MGE will spread resistance for resistance genes present on the element, but not spread sensitivity when resistance genes are absent. Resistances are therefore more likely to be inherited together than resistance and sensitivity. However, we would still expect recombination and mutation to eventually eliminate resistance determinants which do not confer a fitness advantage, even from mobile genetic elements. For example, in the PMEN1 pneumococcal lineage, there is evidence for loss of aminoglycoside resistance from the Tn*916* transposon which encodes tetracycline, and sometimes macrolide, resistance [7]. However, the timescale at which this loss would occur is unclear and there are examples of (presumably) non-advantageous resistance determinants persisting for long time periods [8, 9]. |
| Correlated drug exposure of individual host | Correlated drug exposure at the individual patient level can arise through use of combination therapy, sequential drug exposure (due to treatment failure with the first drug or prophylactic therapy involving antibiotic cycling), or antibiotic exposure among certain patients being particularly high due to co-morbidities. While combination therapy is rare (in the UK for example, monotherapy accounts for 98% of primary care prescriptions [10]), the other two mechanisms play a substantial role in shaping prescription patterns: prophylaxis and repeat prescriptions make up 31% of primary care prescriptions in England [11] and under 10% of patients account for 50% of adult antibiotic prescriptions [12]. It is unclear, however, whether correlation in antibiotic exposure at the individual level can drive selection for MDR: in absence of assortative mixing between patients, the distribution of antibiotic consumption within a population has little effect, although this result may be sensitive to assumptions about the ecology of the bacteria in question (S1 Text Section 5). |

(*Continued*)

**Table 1.** (Continued)

| Process | Notes |
|---|---|
| Resistance status/risk informs antibiotic choice | Prescription practices may also contribute to MDR over-representation in another way: the resistance status of an infection or the presence of risk factors for resistance (e.g. travel to certain areas) affect which antibiotic is prescribed. Strains resistant to a particular antibiotic therefore have higher rates of exposure to other antibiotics. The extent to which this mechanism plays a role likely depends on the type of pathogen: for mostly asymptomatic pathogens, the majority of antibiotic exposure arises from prescriptions due to infections with some other pathogen [13], and the resistance status of this pathogen would therefore not affect the choice of antibiotic. |
| Cost epistasis (lower than expected fitness cost when multiple resistance determinants are present) | There is evidence of cost epistasis between resistance determinants occurring in laboratory competition experiments for some antibiotics (e.g. between streptomycin and rifampicin resistance in *Pseudomonas aeruginosa* [14] and in *E. coli* [15]; streptomycin and nalidixic acid resistance in *E. coli* [16]; and rifampicin and ofloxacin resistance in *Mycobacterium smegmatis* [17]). Furthermore, for plasmid-associated resistance genes, cost epistasis could also arise if the presence of the plasmid in itself incurs a significant fitness cost (rather than the fitness cost depending on the specific resistance genes it carries). However, the extent to which epistasis plays a role *in vivo* remains unclear [18]. In particular, we would not, *a priori*, expect to observe cost epistasis between resistance to antibiotics operating through entirely different mechanisms (e.g. antibiotics targeting protein synthesis and antibiotics targeting the cell wall). |

https://doi.org/10.1371/journal.ppat.1007763.t001

Disease Dynamics, Economics and Policy, available at https://resistancemap.cddep.org/AntibioticResistance.php). For further review of evidence for stable coexistence, see references [19, 20].

Recent work has explored the role of i) host population structure [21–23], ii) pathogen strain structure [20, 21] and iii) within-host dynamics [24] in maintaining the coexistence of antibiotic sensitivity and resistance. In this paper, we identify a structural similarity in the first two categories of model. In these models, coexistence arises through a combination of two factors. First, the presence of groups within the host or pathogen population in which the evolutionary dynamics of resistance are approximately independent from the other groups. Second, the presence of variation in the benefit gained from resistance between these groups, so that antibiotic resistance is selected for in some groups while sensitivity is selected for in others. We show that if this variation is correlated for different antibiotics, models with this structure also predict high levels of association between resistance to different antibiotics: all resistance determinants will tend to be found where the fitness benefit gained from resistance is the greatest. The observed high frequency of multi-drug resistance is therefore in line with ecologically plausible models of coexistence, making these models a parsimonious explanation for both trends.

## Results

### Heterogeneity in the fitness effect of resistance: A generalised model of coexistence

In this section, we discuss competitive exclusion and previously proposed coexistence mechanisms in the context of multidrug resistance. We identify a structural similarity in plausible

models of coexistence [20–22] and show that, in a multidrug context, models with this structure predict MDR over-representation. The model we present captures the dynamics of a bacterial species which is mostly carried asymptomatically (e.g. *E. coli*, *S. aureus* or *S. pneumoniae*), so the probability of a host being exposed to antibiotics does not depend on whether the host is infected with the pathogen [13]. Key results, however, are also applicable when this is not the case (see Discussion).

**Competitive exclusion in single and multi-drug systems.** The coexistence of antibiotic sensitivity and resistance has been previously discussed in the context of competition between two strains (sensitive and resistance strains [19–21] or two resistant strains with different resistance profiles [22]). Simple models of such competition predict competitive exclusion [19]; we start by briefly re-introducing this result and then demonstrate that competitive exclusion also applies in a multidrug context.

We consider a SIS (susceptible-infectious-susceptible) model of resistant and sensitive variants of an otherwise genetically homogeneous pathogen (one strain) circulating in a homogeneous host population. To avoid ambiguity later in the paper, we will refer to the sensitive and resistant variants as 'sub-strains'. Uninfected hosts ($U$) become infected with the resistant ($I_r$) or the sensitive ($I_s$) sub-strain at rate $\beta_r$ and $\beta_s$; infections are cleared at rate $\mu_r$ and $\mu_s$; the sensitive sub-strain experiences an additional clearance rate $\tau$ corresponding to the population antibiotic consumption rate (we assume immediate clearance following antibiotic exposure); and resistance is associated with a fitness cost affecting transmission ($\beta_r = \beta_s c_\beta$, where fitness cost is $1 - c_\beta$ and $0 \leq c_\beta \leq 1$) and/or clearance ($\mu_r = \frac{\mu_s}{c_\mu}$, where fitness cost is $1 - c_\mu$ and $0 \leq c_\mu \leq 1$). The dynamics of this model are described by:

$$\frac{dI_s}{dt} = \beta_s I_s U - (\tau + \mu_s) I_s$$
$$\frac{dI_r}{dt} = \beta_r I_r U - \mu_r I_r \quad (1)$$

This system allows an equilibrium solution where both $I_s$ and $I_r$ are non-zero (i.e. stable coexistence of sensitivity and resistance) only when $\frac{\beta_s}{\tau+\mu_s} = \frac{\beta_r}{\mu_r}$. In other words, the resistant and sensitive sub-strains coexist only when their basic reproductive numbers (the average number of new infections an infected host gives rise to in a fully susceptible population) are equal. When this is not the case, the model predicts competitive exclusion: when resistance provides a fitness advantage ($\frac{\beta_s}{\tau+\mu_s} < \frac{\beta_r}{\mu_r}$), only the resistant sub-strain will be observed, and vice-versa when the sensitive sub-strain is fitter than the resistant sub-strain ($\frac{\beta_s}{\tau+\mu_s} > \frac{\beta_r}{\mu_r}$). Defining $c = c_\mu c_\beta$ and strain clearance rate as $\mu = \mu_s$ to simplify notation, this threshold can be expressed as resistance being selected for when:

$$c[1 + \frac{\tau}{\mu}] > 1 \quad (2)$$

Thus, as reported previously [20], the fitness effect of resistance, which determines whether the resistant sub-strain out-competes the sensitive sub-strain, depends on the population antibiotic consumption rate, the fitness cost of resistance and the strain's mean duration of carriage ($\frac{1}{\mu}$), because longer carriage episodes have a greater risk of antibiotic exposure than shorter carriage episodes [20].

We now extend this model to $n$ antibiotics administered as monotherapy (we do not model combination therapy—see S1 Text Section 5). Sub-strains can be either sensitive or resistant to each antibiotic, giving a total of $2^n$ competing sub-strains. Similarly to the single drug model

presented above, resistance to each antibiotic $j$ has a transmission associated fitness cost $1 - c_{\beta j}$ and/or a clearance associated fitness cost $1 - c_{\mu j}$. We assume no cost epistasis between resistance determinants: the fitness cost of resistance to an antibiotic does not depend on which other resistances are present on the sub-strain. Note that this assumption is not necessary for the demonstration of competitive exclusion in a multidrug context, but becomes important in later sections of paper. For consistency, we introduce it here. Therefore, a sub-strain $k$, resistant to the set of antibiotics $R_k$, has transmission rate $\beta_k = \beta \prod_{j \in R_k} c_{\beta j}$ and clearance rate $\mu_k = \frac{\mu}{\prod_{j \in R_k} c_{\mu j}}$ where $\beta$ and $\mu$ are the transmission and clearance rates of the fully sensitive sub-strain. In addition, each sub-strain is cleared by the antibiotics it is sensitive to (set of antibiotics $S_k$), giving a total clearance rate of $\lambda_k = \mu_k + \sum_{j \in S_k} \tau_j$, where $\tau_j$ is the consumption rate of antibiotic $j$. The dynamics of each sub-strain are therefore described by:

$$\frac{dI_k}{dt} = \beta_k I_k U - \lambda_k I_k \tag{3}$$

for all $k \in \{1, \ldots, 2^n\}$ and with $U = 1 - \sum_k I_k$. At equilibrium, $\frac{\lambda_k}{\beta_k} = U$ holds for all strains with non-zero frequency—we therefore recover the result from the single drug model: sub-strains can only coexist when they have the same reproductive number. When this is not the case, the frequency of resistance to each antibiotic is either 0% or 100% and a single resistance profile with the highest reproductive number $\frac{\beta_k}{\lambda_k}$ is expected to out-compete all others. In a multidrug context, therefore, coexistence-maintaining mechanisms are necessary to explain why multiple different resistance profiles are observed.

**Coexistence through heterogeneity in the fitness effect of resistance: Single drug context.** In this section, we note a structural similarity in plausible models of coexistence: a number of recently proposed coexistence mechanisms work by introducing variation in the fitness effect of resistance within either the host population or the pathogen population. We show that models with this structure can be simplified to a series of independent SIS models, which will allow us to gain insight into the pattern of association between resistance to different antibiotics in a multidrug context. We start by presenting a simple model for conceptual insights; additional complexity is explored in later sections.

In the first class of models we consider, coexistence arises from host population structure: assortatively mixing groups within the host population promote coexistence if the groups differ in the fitness effect of resistance, thus creating niches for resistance and sensitivity within the host population. Sources of population structure that have been proposed to promote coexistence in this manner include hospital vs community settings [22] (where variation in the fitness benefit of resistance would arise from variation in antibiotic consumption rate) and age groups [20, 21, 23] (where variation in the fitness benefit of resistance would arise from variation in both antibiotic consumption rate and clearance rate). Other potentially relevant host groups include geographic areas and socio-economic groups.

For assortative mixing between host groups to promote coexistence, transmission between groups must be very low: even modest transmission between groups causes the groups to act as a single population and therefore abolishes coexistence [23] (see also section *Extension: additional complexity*). By treating this very low transmission as no transmission, we can represent host structure by modelling each of the host groups as a separate SIS model, with dynamics captured by Eq (1). In other words, we model competition between sensitivity and resistance within each host group as independent of the other host groups (Fig 1).

In the second class of models, coexistence arises from heterogeneity within the pathogen, rather than host, population: the heterogeneity in fitness effect of resistance arises from the

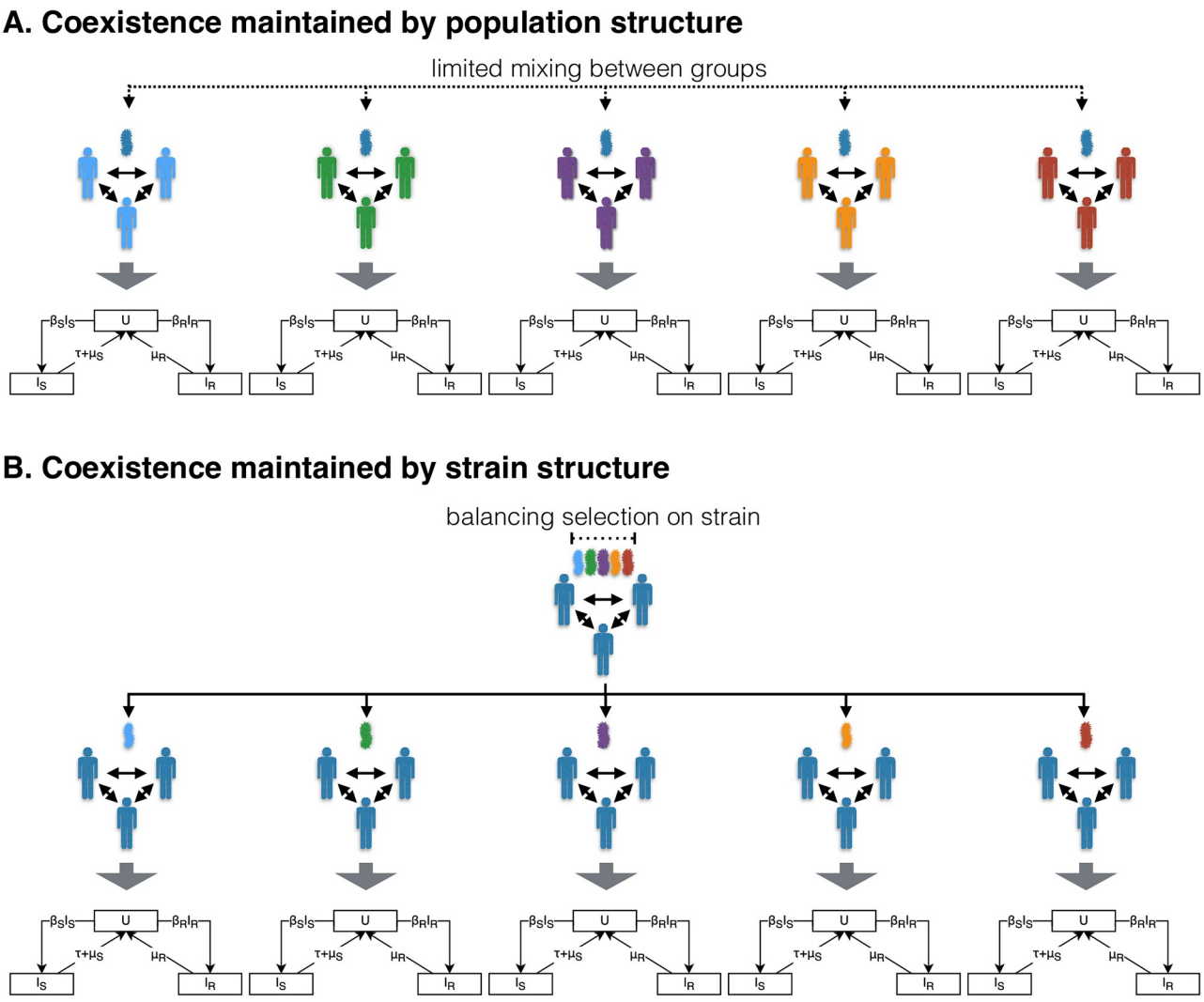



**Fig 1. Host population and strain structure in maintaining coexistence of antibiotic sensitivity and resistance.** Illustration of how host population (panel A) and strain (panel B) structure maintain coexistence by introducing heterogeneity in the fitness effect of resistance and thus creating niches for sensitivity and resistance within the population. Each of the SIS model diagrams represents the resistance dynamics described by Eq (1). A: The resistance dynamics of assortatively mixing host groups can be modelled as independent SIS models by assuming no transmission between groups. Heterogeneity in the fitness effect of resistance arises from between host group differences in antibiotic consumption rate or clearance rate. B: The resistance dynamics of pathogen strains maintained by balancing selection can be modelled as independent SIS models by assuming no recombination. Heterogeneity in the fitness effect of resistance arises from between strain differences in mean duration of carriage (i.e reciprocal of clearance rate).

https://doi.org/10.1371/journal.ppat.1007763.g001

presence of strains with different durations of carriage, maintained by balancing selection on the duration of carriage locus (e.g. serotype-specific acquired immunity allowing coexistence of serotypes with different durations of carriage in the pneumococcus [25]) [20]. Hence, similarly to assortatively mixing host groups in the first class of models, strains with different durations of carriage act as niches for sensitivity and resistance: coexistence is maintained by competition between sensitivity and resistance occurring independently within each strain. By assuming no recombination, we can again represent competition between sensitivity and resistance within each strain as a separate SIS model (Fig 1). Note that this class of models requires the presence of balancing selection maintaining diversity at the duration of carriage locus. In the simplified representation, this balancing selection is not modelled explicitly—coexistence

of the strains differing in duration of carriage is assumed (see S1 Text Section 1 for further discussion of this point).

Thus models in which coexistence arises from heterogeneity in the fitness effect of resistance—either within the host or pathogen population—can be represented by a series of independent SIS models. We refer to these individual SIS models as strata. In the case of a single strain circulating in a structured host population, the strata correspond to assortatively mixing host groups (e.g. age classes). In the case of multiple strains circulating in a homogeneous host population, the strata correspond to different strains (e.g. serotypes in the pneumococcus). When both strain and population structure are present, each stratum corresponds to a particular strain circulating in a particular host group. Following from Eq (2), resistance out-competes sensitivity in stratum *pi* (host group *p* and strain *i*) when:

$$c(1+\frac{\tau_p}{\mu_{pi}})>1. \tag{4}$$

**Coexistence through heterogeneity in the fitness effect of resistance: Multidrug context.** We now extend this model to multiple antibiotics, which may differ in fitness cost and consumption rate. We assume no cost epistasis between resistance determinants: resistance to antibiotic *a* has the same fitness cost (i.e. the same $c_a$) in presence and absence of resistance to antibiotic *b*. We also assume that different antibiotics are consumed in the same proportions in all host groups: antibiotic *a* accounts for proportion $\gamma_a$ of total antibiotic consumption, with $\sum_{a=1}^{n}\gamma_a=1$, where *n* is the number of different antibiotics. The consumption rate of antibiotic *a* for host group *p* is therefore $\gamma_a\mathcal{T}_p$, where $\mathcal{T}_p$ is the total antibiotic consumption rate of group *p*.

Under these assumptions, following from Eq (2), resistance to antibiotic *a* out-competes sensitivity in host group *p* and strain *i* when:

$$c_a[1+\frac{\gamma_a\mathcal{T}_p}{\mu_{pi}}]>1 \tag{5}$$

As before, there is no coexistence within the strata: below this threshold, sensitivity out-competes resistance. In a multi-drug context, a single resistance profile will therefore out-compete all others within each stratum.

**Heterogeneity in the fitness effect of resistance: Predicted patterns of resistance.** We can separate Eq (5) into stratum (i.e host group and pathogen strain) and antibiotic related effects (left-hand and right-hand sides of Inequality 6, respectively):

$$\frac{\mathcal{T}_p}{\mu_{pi}}>\frac{1}{\gamma_a}(\frac{1}{c_a}-1) \tag{6}$$

We call the ratio $\frac{\mathcal{T}_p}{\mu_{pi}}$ resistance proneness ($P_{pi}$) and the ratio $\frac{1}{\gamma_a}(\frac{1}{c_a}-1)$ resistance threshold ($T_a$). $P_{pi}$ reflects how advantageous resistance is within stratum *pi*. High antibiotic consumption (high $\tau_p$) and low clearance rate (low $\mu_{pi}$) lead to high resistance proneness. $T_a$ reflects how advantageous resistance against antibiotic *a* needs to be for it to be selected for. High fitness cost (low $c_a$) and making up a low proportion of total antibiotic consumption (low $\gamma_a$) lead to high resistance threshold. Rewriting Eq (6) using this notation, resistance to antibiotic *a* is selected for in stratum *pi* when:

$$P_{pi}>T_a \tag{7}$$

Resistance proneness depends only on stratum and resistance threshold depends only on antibiotic. As a consequence, the ordering of strata by resistance proneness is independent of

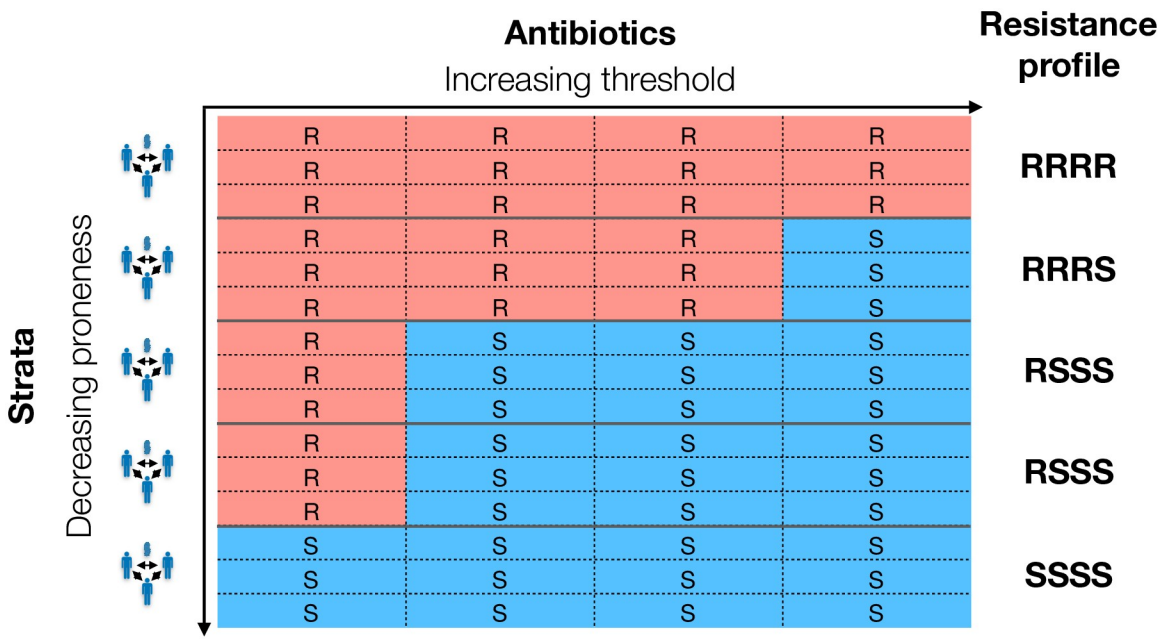



**Fig 2. Example of a set of resistance profiles from a system with five strata and four antibiotics.** Each row in the table corresponds to the resistance profile of one isolate—i.e. there are three isolates from each strata (equal sampling/size of strata is not necessary). Competitive exclusion within a stratum means all isolates from one stratum have the same resistance profile. The strata have been arranged from top to bottom in order of decreasing resistance proneness ($P_{pi} = \frac{T_p}{\mu_{pi}}$). The antibiotics have been arranged left to right in order of increasing resistance threshold ($T_a = \frac{1}{\gamma_a}(\frac{1}{c_a} - 1)$), or, equivalently, decreasing resistance frequency. Resistance to a particular antibiotic outcompetes sensitivity in a stratum when the resistance proneness of the stratum is greater than the resistance threshold of the antibiotic. Resistance proneness being independent of antibiotic and resistance threshold being independent of stratum leads to nested resistance profiles (i.e. rarer resistances only observed in the presence of more common ones) and complete linkage disequilibrium between resistances. See Fig A in S1 Text for an example of a set of non-nested resistance profiles.

https://doi.org/10.1371/journal.ppat.1007763.g002

antibiotic and the ordering of antibiotics by resistance threshold is independent of stratum. Therefore, for a set of $m$ strata, with resistance proneness $P_1 < P_2 < \ldots < P_m$ and for a set of $n$ antibiotics, with resistance thresholds $T_1 < T_2 < \ldots < T_n$, the following will hold:

$$P_i > T_a \Rightarrow P_i > T_b \ \forall \ b < a \tag{8}$$

$$P_i < T_a \Rightarrow P_i < T_b \ \forall \ b > a \tag{9}$$

That is, if resistance to antibiotic $a$ is selected, resistance to antibiotics with a lower resistance threshold will also be selected for. Conversely, if resistance to $a$ is not selected for, resistance to antibiotics with a higher resistance threshold will also not be selected for. Thus, the ordering of antibiotics by resistance threshold ($T_1 < T_2 < \ldots < T_n$) determines the ordering of antibiotics by resistance frequency: the higher the resistance threshold, the lower the resistance frequency and resistance to a particular antibiotic will only be seen on resistance profiles with all more frequent resistances. Therefore, the fitness variation model predicts non-zero frequencies for only $n + 1$ out of the $2^n$ possible resistance profiles: the only profile with resistance multiplicity of $m$ (i.e. resistance to $m$ antibiotics) will be the one with the $m$ most common resistances (Fig 2). This pattern of resistance is referred to as 'nested' and predicts strong association between resistance to different antibiotics, with all resistance pairs in complete linkage disequilibrium (D' = 1, where D' is the normalised coefficient of linkage disequilibrium (LD)—see Methods).

## Extension: Additional complexity

In this section, we explore how introducing additional complexity to the simplified model affects our predictions about association between resistance determinants and nestedness.

**The effect of resistance on clearance rate.** The model presented above does not take into account the effect of the absence/presence of resistance to other antibiotics on the resistance threshold of antibiotic *a*. This is a simplification because i) in the absence of resistance against other antibiotics, the exposure to these antibiotics will contribute to clearance and ii) the presence of resistance to other antibiotics will affect clearance if the fitness cost of resistance increases clearance rate. In a multidrug context therefore, Eq (5) is an approximation.

In S1 Text Section 3, we show that this approximation does not meaningfully affect our predictions about association between resistance determinants and nestedness of resistance profiles. The effect of resistance on clearance rate does not give rise to incomplete linkage disequilibrium ($D' < 1$) and non-nested resistance profiles, except under very specific circumstances: if the fitness cost of resistance affects clearance rate and more commonly prescribed antibiotics also have higher fitness cost. Even when incomplete linkage disequilibrium is possible theoretically, the parameter range under which it arises is extremely narrow (see S1 Text Section 3), suggesting incomplete linkage disequilibrium and non-nested resistance profiles being observed because of the effect of resistance on clearance rate is unlikely.

**Intergroup transmission and recombination.** In the model presented above, strata are fully independent, with no transmission between host groups and no recombination. As discussed, this is a simplification: coexistence can be maintained in the presence of mixing between strata if the rate of mixing is low enough [23]. In order to investigate the effect that mixing between strata has on our predictions about the association between resistance determinants, we construct i) a model with three different antibiotics and five host groups differing in clearance rate with transmission allowed between host groups and ii) a model with three different antibiotics and five strains differing in clearance rate with recombination allowed at the duration of carriage locus (see Methods for details). In both models, complete linkage disequilibrium is maintained in the presence of mixing between strata (Fig 3): mixing does not introduce any source of selection that would favour non-nested resistance profiles.

We also investigate the effect of recombination at the resistance loci in the strain structured model (see Methods). Unlike recombination at the duration locus, recombination at the resistance loci breaks up linkage disequilibrium (Fig 3), decreasing the magnitude of association between resistance determinants and giving rise to maladapted allele combinations ('recombination load'). However, this effect is gradual and high levels of LD are maintained even at unrealistically high rates of recombination (see S1 Text Section 4). We do not implement recombination in the population structured model: we assume recombination requires co-infection and expect no co-infection in this model because of competitive exclusion within each host group. We expect that recombination together with inter-group transmission would similarly give rise to decreased LD and recombination load.

**Imperfectly correlated strata.** In the simple model we present, the prediction of complete LD arises because we can separate the variation in the fitness effect of resistance into strata-related (i.e. pathogen and host) and antibiotic-related effects. This separability means that resistance proneness of a stratum is independent of antibiotic, which gives rise to complete LD between resistance to different antibiotics and resistance profiles with nested structure. This separability requires two assumptions: first, that the fitness cost of a particular resistance is the same in all strata (i.e. no variation in the fitness cost of resistance between strains) and second, that different types of antibiotics are consumed in the same proportions in all strata (i.e.

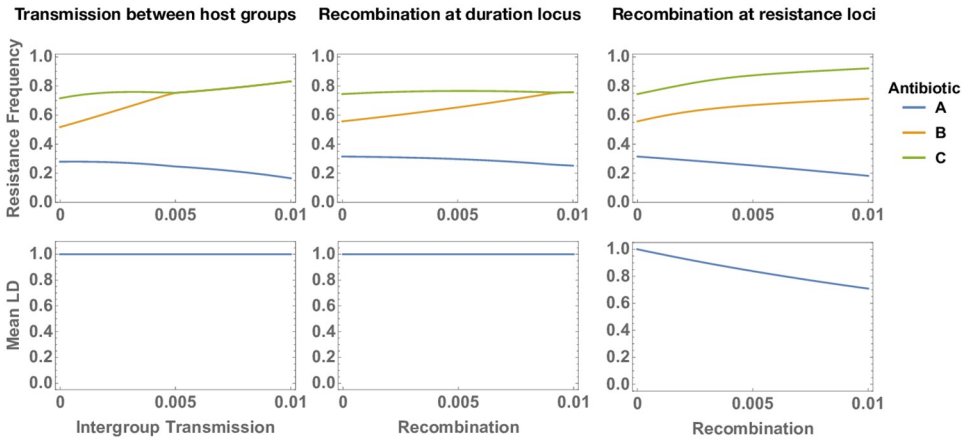



**Fig 3. Strain frequencies and mean linkage disequilibrium (LD) between resistances in three models with three antibiotics (*A*, *B* and *C*) consumed at different rates.** Left: a model with host population structure (five assortatively mixing host groups) with increasing levels of intergroup transmission. The rate of intergroup transmission on the x-axis (parameter $m$ in the model represented by Eq (10), see Methods) reflects the proportion of transmission events that occur between, instead of within, host group. Middle: a model with strain structure (five strains differing in duration of carriage) with increasing rates of recombination at the duration of carriage locus. Recombination rate on the x-axis (parameter $r$ in the model represented by Eq (11), see Methods) reflects the probability of co-infection, the probability of recombination occurring during co-infection and the probability of the recombinant strain being transmitted. Right: the same model with strain structure (five strains differing in duration of carriage) with increasing rates of recombination at the resistance loci.

https://doi.org/10.1371/journal.ppat.1007763.g003

variation in the *rate* at which host groups consume antibiotics, but not in the mixture of antibiotic types).

Both of these requirements may represent oversimplifications of resistance dynamics. In the case of the first assumption, there is no direct evidence for stable variation in the fitness cost of resistance (i.e. variation in fitness cost maintained by balancing selection, preventing the lower fitness cost phenotype reaching fixation). However, the processes determining fitness cost are not fully understood and the fitness cost of resistance mutations is thought to depend on both genetic background and environment [26]. It is therefore difficult to rule out that stable between-strain variation might exist.

The second assumption (correlated antibiotic prescription profiles) may be plausible for some host groups—for example, if strata correspond to geographical areas of the same country. However, this assumption does not hold for other host groups. Children and adults have different antibiotic consumption profiles for some antibiotics: for example, fluoroquinolones are primarily used in adults but not children [27], which has been proposed as an explanation for why association between resistances is weaker for fluoroquinolones than for other antibiotics [2].

To test the effect of allowing antibiotic consumption profiles to differ between strata, we construct a two drug model with ten assortatively mixing host groups (see Methods for details). For each drug, five of these host groups consume antibiotics at a high rate (which selects for resistance) and five consume antibiotics at a low rate (which selects for sensitivity). When the consumption rates of the two antibiotics are perfectly correlated across the host groups, the resistance are in complete linkage disequilibrium ($D' = 1$). This linkage disequilibrium decreases with decreasing correlation in the relative consumption of the two antibiotics across groups (Fig 4). The extent of this decrease depends on whether the resistance proneness of the strata is entirely determined by the antibiotic consumption rate or whether the strata also differ in clearance rate. In the presence of large variation in clearance rate, even negatively correlated antibiotic consumption rates can give rise to positive LD (Fig 4).

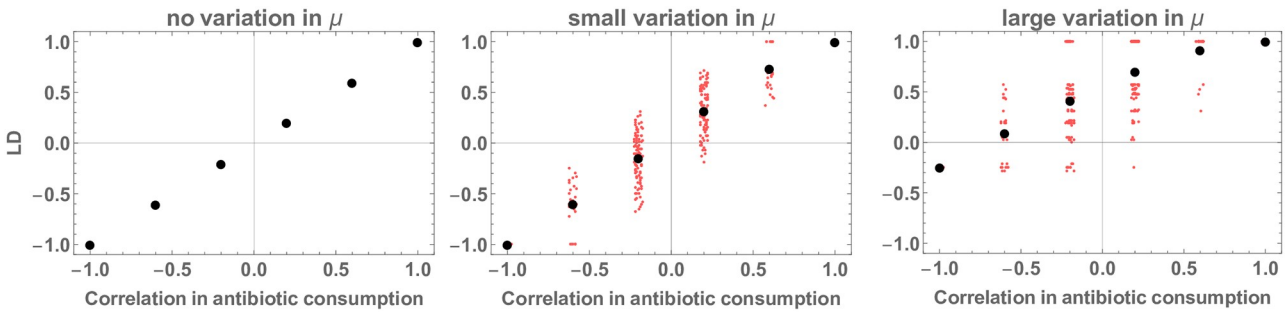



**Fig 4. Linkage disequilibrium (measured as $D'$) between resistances in a two drug model with ten assortatively mixing host groups, as a function of the Spearman correlation between antibiotic consumption rates for the two antibiotics across the ten host groups.** For each drug, five of these host groups consume antibiotics at a high rate ($\tau_{high} = 0.075$ which selects for resistance when $\mu = 1$) and five consume antibiotics at a low rate ($\tau_{low} = 0.025$ which selects for sensitivity when $\mu = 1$). Consumption rates vary from perfectly anticorrelated (all host groups consuming the first antibiotic at high rate consume the second antibiotic at low rate) to perfectly correlated (all host groups consuming the first antibiotic at a high rate also consume the second antibiotic at a high rate). Left-hand panel: all host groups have the same clearance rate ($\mu = 1$). Middle panel: small variation in clearance rate between host groups ($0.5 \leq \mu \leq 1.5$). Each red marker corresponds to one possible configuration of antibiotic consumption and clearance rates (see Methods for details), the black markers represent the average of all possible configurations with the same correlation between antibiotic consumption rates (horizontal jitter is for visualisation purposes only). Right-hand panel: similar to middle panel but with larger variation in clearance rate between host groups ($0.25 \leq \mu \leq 2$). Other parameters are $\beta = 2$, $c_\beta = 0.95$ for both antibiotics, $c_\mu = 1$ for both antibiotics.

https://doi.org/10.1371/journal.ppat.1007763.g004

## Model predictions are consistent with trends in bacterial datasets

**High levels of association between resistance determinants.** The model we present is motivated by the observed over-representation of MDR. Previous work [2] has quantified this over-representation in terms of pairwise coefficients of determination ($r^2$) between resistances. This captures the extent to which isolates that are resistant to one antibiotic are also resistant to another antibiotic. The prediction made by the fitness variation model is slightly different: it predicts high $D'$ between resistances and a high proportion of resistance profiles with a nested structure. $D'$ captures the extent to which resistance to two antibiotics will be found in the same isolates, given the observed resistance frequencies. The key difference is that while $r^2$ is affected by how similar resistance frequencies are and by the distribution of the resistance determinants, $D'$ is only affected by the latter.

We measure $D'$ and the proportion of nested antibiograms in six bacterial datasets for which data on resistance to multiple antibiotics was available (four hospital datasets from the United States for different species and two pneumococcal datasets from Massachusetts and Maela—see Methods for additional details). A high proportion of antibiograms have a nested structure (Fig 5). Values of $D'$ for specific antibiotic pairs range from very low to very high, but average values are always positive and generally above 0.5 (Table 2 and Fig G in S1 Text). In the instances where the minimum $D'$ value is negative or close to zero, at least one of the resistances in the pair this $D'$ value corresponds to was present at very low frequency in the dataset (0.08% for the *E. coli* dataset, 1.3% in the *S. aureus* dataset, 6.1% for the Maela pneumococcal dataset).

The trends in Table 2 are broadly consistent with the predictions of our model (i.e. high $D'$ and high proportion of nested antibiograms). We do not address the question of whether these trends are more consistent with the fitness variation model than other possible mechanisms of MDR over-representation (Table 1) because it is unclear what patterns of $D'$ and nestedness these mechanisms predict (see Discussion).

**Duration of carriage predicts resistance multiplicity.** The fitness variation model predicts that duration of carriage and antibiotic consumption rate within strata will determine resistance multiplicity. Fully testing this prediction is challenging, because we do not have a

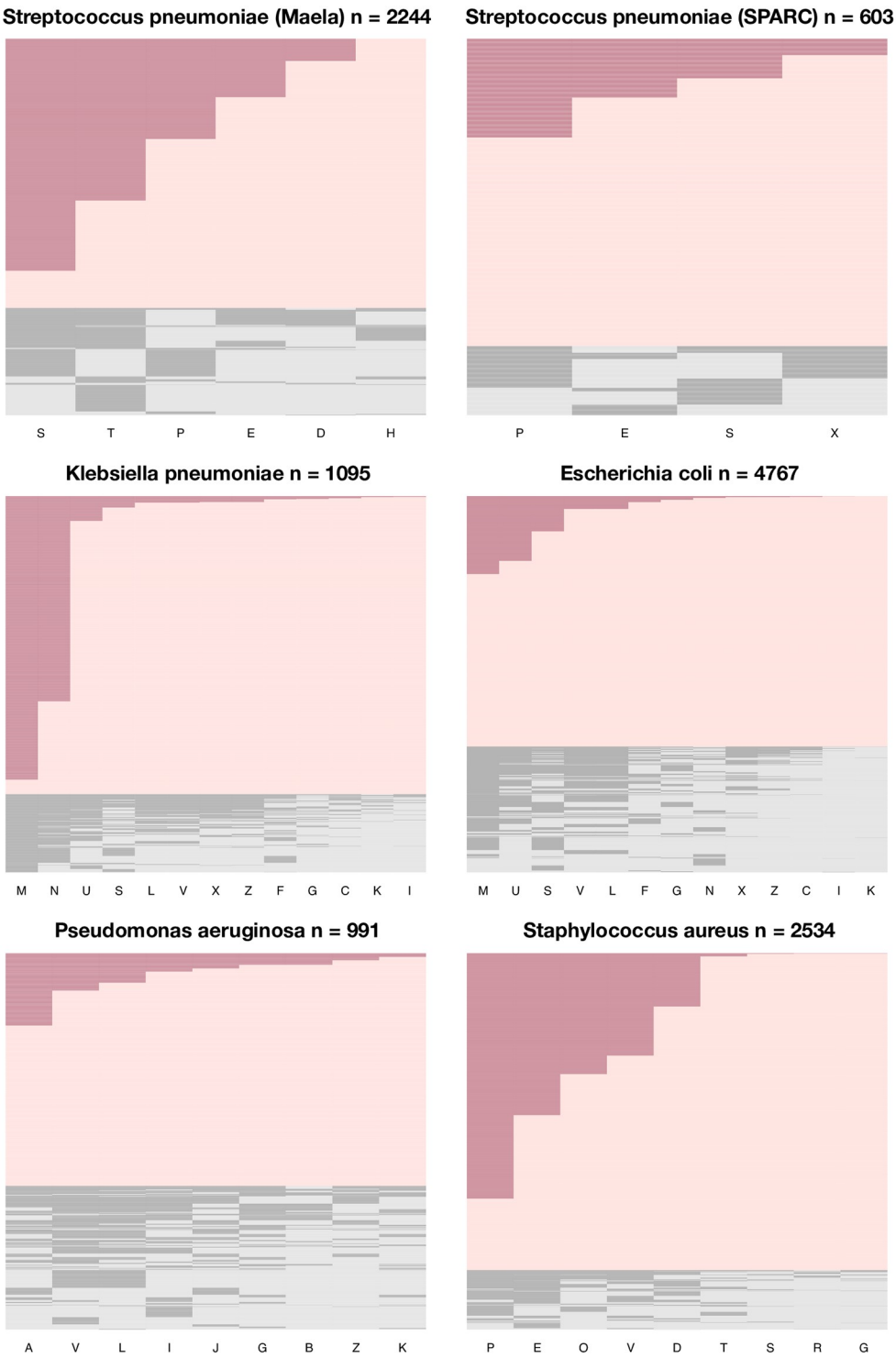



**Fig 5. Antibiograms (resistance profiles) in the six bacterial datasets.** Dark shading indicates resistance, light shading indicates sensitivity. Antibiograms with a nested structure are coloured red. Each row represents the antibiogram of one isolate (sorted by nestedness and resistance multiplicity). Columns represent antibiotics, ordered by frequency of resistance. A: aztreonam, B: tobramycin, C: cefepime, D: clindamycin, E: erythromycin, F: cefoxitin, G: gentamicin, H: chloramphenicol, I: imipenem, J: piperacillin, K: amikacin, L: ciprofloxacin, M: ampicillin, N: nitrofurantoin, O: oxacillin, P: penicillin, R: rifampin, S: trimethoprim-sulfamethoxazole, T: tetracycline, U: ampicillin-sulbactam, V: levofloxacin, X: ceftriaxone, Z: ceftazidime.

https://doi.org/10.1371/journal.ppat.1007763.g005

**Table 2. Mean pairwise LD between antibiotic pairs ($\hat{D}'$) and proportion of resistance profiles that are nested for six bacterial datasets.**

| Species | Setting | n | Drugs | $\hat{D}'$ (min,max) | Nested |
|---|---|---|---|---|---|
| *Pseudomonas aeruginosa* | Hospital (USA) | 991 | 9 | 0.55 (0.22,0.95) | 0.62 |
| *Escherichia coli* | Hospital (USA) | 4767 | 13 | 0.54 (-1,1) | 0.67 |
| *Klebsiella pneumoniae* | Hospital (USA) | 1095 | 13 | 0.75 (0.24,1) | 0.79 |
| *Staphylococcus aureus* | Hospital (USA) | 2534 | 9 | 0.58 (0.02, 1) | 0.84 |
| *Streptococcus pneumoniae* | Community (Maela) | 2244 | 6 | 0.46 (-0.66,0.95) | 0.71 |
| *Streptococcus pneumoniae* | Community (Massachusetts) | 603 | 4 | 0.58 (0.32,1) | 0.82 |

The *n* column indicates the number of isolates in the dataset and the *drugs* column indicates the number of drugs resistance was tested for. Positive $D'$ indicates resistance determinants tend to appear together. The values in parentheses give the range of pairwise $D'$ in the dataset. See Methods for details of datasets.

https://doi.org/10.1371/journal.ppat.1007763.t002

full understanding of which host and pathogen characteristics are relevant in defining the strata. To partially test the prediction, we test the association between duration of carriage and resistance using a dataset of pneumococcal carriage episodes and associated durations of carriage [28] ('Maela dataset', see Methods). The average resistance multiplicity of a serotype is indeed positively associated with the serotype's average duration of carriage (Kendall rank correlation 0.27 95% CI 0.05-0.46, n = 38 excluding serotypes with fewer than 10 observations). A caveat here is that the direction of causality for this association is not entirely clear: we suggest long duration of carriage selects for resistance but resistance would also be expected to lead to longer duration of carriage through decreased clearance from antibiotic exposure. However, at the serotype level, differences in duration of carriage are thought to arise from the properties of serotype capsules [29] (rather than differences in antibiotic resistance). This suggests that longer duration of carriage favouring resistance does indeed contribute to the association between duration of carriage and resistance multiplicity.

## Discussion

### Generalised model of coexistence predicts high frequencies of multidrug resistance

In this paper, we approach the question of explaining observed patterns of association between resistance to different antibiotics ('MDR over-representation') in terms of understanding the competition between strains with different resistance profiles. We consider recent models of the coexistence of antibiotic sensitive and antibiotic resistant strains [20–23] in which coexistence is maintained by heterogeneity in the fitness effect of resistance, arising either from heterogeneity in the rate of antibiotic consumption and/or difference in duration of carriage. We present a generalised version of these types of models, in which competition between antibiotic sensitivity and resistance is simplified to a series of independent sub-models (strata). We show that this model structure also gives rise to MDR over-representation because resistance to all antibiotics will be selected for in the strata where the fitness benefit of resistance ('resistance proneness') is the highest. Therefore, our results suggest that two pervasive trends in resistance dynamics, the robust coexistence of antibiotic sensitive and resistant strains and the over-representation of multidrug resistance, can both be explained by heterogeneity in the fitness effect of resistance within the host or pathogen population.

We first present a simplified model for conceptual insights and then explore how additional complexity affects predicted trends. Under the strong assumption of identical antibiotic prescription patterns in all strata and no recombination, this model predicts complete linkage disequilibrium ($D' = 1$) between resistance to all antibiotics. Relaxing these assumption decreases

the magnitude of linkage disequilibrium, giving rise to values of $D'$ similar to those observed in multiple bacterial datasets. High $D'$ is maintained even at unrealistically high recombination rates. A lower correlation in antibiotic consumption profiles between strata leads to lower values of $D'$. However, the effect is gradual and the magnitude of the decrease depends on whether the strata also differ in clearance rate. Thus, even in context where patterns of prescription differ considerably between host groups, we would still expect a degree of association between resistance determinants when variation in duration of carriage contributes to variation in the fitness effect of resistance.

Although the model builds on work exploring the stable coexistence of antibiotic sensitivity and resistance and coexistence is robustly observed in multiple datasets, the prediction that variation in the fitness effect of resistance leads to MDR over-representation does not require coexistence to be stable. We would expect MDR over-representation in the presence of fitness variation, even when this variation is not enough to maintain stable coexistence: for all antibiotics, the increase of resistance frequencies towards fixation would occur most rapidly in the populations with the greatest selection pressure for resistance. Under these circumstances, fitness variation would give rise to transient MDR over-representation.

## Prevalence and sources of fitness variation

Our results show that when variation in the fitness effect of resistance is present and when this variation is at least partially correlated for different antibiotics, it will give rise to MDR over-representation. The extent to which this mechanism accounts for observed patterns of MDR over-representation therefore depends on the extent to which this type of fitness variation is present in pathogen populations.

It is not entirely straightforward to evaluate how common variation in the fitness effect of resistance is. Wide-spread coexistence of sensitivity and resistance is not direct evidence for the pervasiveness of fitness variation because coexistence may not always arise through this mechanism. Although the majority of mechanisms proposed to date [20–23] work through fitness variation, other mechanisms are also possible [19]. In particular, recent modelling suggests that co-infection with sensitive and resistant strains gives rise to frequency-dependent selection for resistance and thus promotes coexistence [24]. However, the magnitude of this effect depends on the nature of within-host competition [24], for which there is limited data. Thus while theoretically plausible, the extent to which this mechanism contributes in practice is still unclear. It is worth noting that different coexistence mechanisms are not mutually exclusive. If coexistence arises through a combination of fitness variation and other mechanisms, we would *a priori* still expect the fitness variation to give rise to MDR over-representation.

In the work presented here, we consider fitness variation arising from heterogeneity in antibiotic consumption between host groups (hospitals vs communities, geographic regions, age classes) and from heterogeneity in duration of carriage between host groups (age classes) and between strains (pneumococcal serotypes). This is not an exhaustive list of possible sources of heterogeneity. For example, serotype does not fully account for heritable variation in pneumococcal duration of carriage [28], suggesting other genetic traits also play a role in determining carriage duration. In light of recent results suggesting wide-spread negative frequency-dependent selection in bacterial genomes [30, 31], it is not implausible to suggest these duration of carriage loci may also be under frequency-dependent selection. If so, diversity at these loci would create another source of variation in the fitness effect of resistance and hence promote coexistence and MDR over-representation. More broadly, variation in the fitness effect of resistance may arise through different mechanisms for pathogens with a different ecology than modelled in this work. For example, we have modelled a pathogen that is mostly carried

asymptomatically and therefore exposed primarily to antibiotics prescribed against other infections. For pathogens where antibiotics prescribed due to infection with the pathogen itself contribute to a significant proportion of antibiotic exposure, the presence of strains differing in invasiveness would give rise to between-strain variation in antibiotic exposure and heterogeneity in the fitness effect of resistance. For bacterial species able to multiply both in hosts and in the environment, the sort of structure and heterogeneity considered in this work may also arise from differences between environmental niches.

## Unanswered questions

This study does not fully address the role antibiotic prescription patterns in MDR over-representation: we highlight two important remaining questions. Firstly, in the modelling framework used in this study, the distribution of drug consumption within a stratum (i.e. a well-mixed population) does not have an impact on MDR over-representation (S1 Text Section 5). In other words, the presence of host groups consuming antibiotics at different rates only promotes MDR over-representation if there is very little transmission between these host groups: individual-level correlation in antibiotic exposure is not predicted to promote multi-drug resistance. We have not explored this result in detail—it may arise because the model predicts competitive exclusion within a stratum. Secondly, in contrast to the distribution of antibiotic consumption within a stratum, our results suggest that the distribution of antibiotic consumption between strata does matter: the prediction of MDR over-representation is sensitive to how correlated prescription profiles are and the extent of this sensitivity depends on whether variation in duration of carriage is also present. Relating these theoretical results to observed correlations in the antibiotic consumption between different host groups and to the extent of assortative mixing between these groups will provide additional insights into observed patterns of MDR (e.g. why the association between some drugs is higher than others).

The fitness variation model playing a role in MDR over-representation does not preclude a potential role for other mechanisms in contributing to the trend (Table 1). This study does not address the relative extent to which the different possible mechanisms contribute to MDR over-representation. This is for two reasons. Firstly, it is unclear what the patterns of MDR predicted by alternative mechanisms of MDR over-representation are. Secondly, we do not have a full understanding of which host and pathogen characteristics are relevant in defining the strata so it is difficult to directly address whether these traits are predictors of MDR. One alternative strategy for establishing the extent to which the fitness variation model contributes to MDR over-representation would be to assess patterns of association between resistance determinants in a single strain circulating in a well-mixed host population (i.e. a single stratum). The fitness variation model predicts no MDR over-representation (as defined by $D' > 0$) under these circumstances. Therefore, if linkage disequilibrium is observed under these conditions, this would indicate that fitness variation is not the only mechanism of MDR over-representation. Furthermore, the magnitude of linkage disequilibrium could inform the relative contribution of the fitness variation mechanism: observing similar levels of linkage disequilibrium within strata and within the whole population would suggest the fitness variation is not a necessary mechanism for generating MDR over-representation.

## Public health implications

From a public health perspective, the fitness variation model makes two concerning predictions. Firstly, we predict frequencies of pan-resistance will be high: in a perfectly nested set of resistance profiles, the frequency of pan-resistance is equal to the frequency of the rarest resistance. As a consequence, we would expect resistance arising in response to adoption of new

antibiotics or increased usage of existing antibiotics to appear on already multidrug resistant lineages—an observation which has been made for the emergence of ciprofloxacin resistance in *N. gonorrhoeae* in the United States [32].

Secondly, our analysis has implications for the effectiveness of potential interventions against MDR. The variation in the fitness effect of resistance to different antibiotics need not be perfectly correlated for it to promote MDR over-representation. If the variation in fitness effect is maintained by multiple factors (e.g. differential antibiotic consumption between populations and variation in clearance rates), removing one of these factors (e.g. changing patterns of prescription so that consumption of different antibiotics is no longer correlated between host groups) may have limited impact on MDR over-representation.

The fitness variation model provides an explanation for MDR over-representation that is consistent with long term stability in resistance frequencies. This is relevant when considering temporal trends in resistance frequencies and predicting the future burden of resistance: other explanations for MDR over-representation (e.g. cost epistasis, correlated antibiotic exposure at the individual level—see Table 1) often require MDR strains to have an overall fitness advantage over strains with lower resistance multiplicity. This would imply that the higher than expected frequency of MDR is evidence for MDR strains out-competing other strains and thus suggest that MDR strains will eventually take over. Conversely, in the model we present, MDR strains are not out-competing other strains: all resistance frequencies are at equilibrium and MDR over-representation arises from the distribution of resistance determinants. It is worth noting, however, that even in the context of the fitness variation model, on a very long timescale, we might expect the frequency of resistance to rise if bacteria are able to evolve resistance mechanisms that carry a lower fitness cost.

## Conclusion

We show that previously proposed models in which coexistence of antibiotic sensitivity and resistance is maintained by heterogeneity in the fitness effect of resistance also predict high frequencies of multidrug resistance. The pervasive trends of coexistence and MDR over-representation can therefore be considered, at least partially, facets of the same phenomenon. We do not propose that the model we present fully explains observed patterns of association between resistance determinants. However, this effect should be considered when evaluating the role of antibiotic-specific MDR promoting mechanisms. From a public health point of view, the model we present is concerning because it predicts high frequencies of pan-resistance. On the other hand, heterogeneity in the fitness effect of resistance as an explanation for MDR over-representation allows reconciling this trend with long term stability in resistance frequencies.

# Methods

## Datasets

The Maela pneumococcal dataset [33], collected from a refugee camp on the border of Thailand and Myanmar from 2007 to 2010, consisted of 2244 episodes of carriage, with associated antibiograms and carriage durations. Data were obtained from, and durations of carriage calculated by, Lees et al. [28] (S1 File). Data on antibiotic sensitivity was provided for ceftriaxone, chloramphenicol clindamycin, erythromycin, penicillin, co-trimoxazole (trimethoprim/sulfamethoxazole) and tetracycline. Ceftriaxone was excluded from the analysis because data was missing for a large proportion of isolates (44%). The Massachusetts pneumococcal dataset, collected as part of the SPARC (*Streptococcus pneumoniae* Antimicrobial Resistance in Children) project [34], was obtained from Croucher et al. (2013) [35] (data

available from Croucher et al [35]). Croucher et al. reported minimum inhibitory concentrations (MICs) for penicillin, ceftriaxone, trimethprim, erithromycin, tetracycline and chloramphenicol. Tetracycline and chloramphenicol were excluded from the analysis because data was missing for a large proportion of isolates (47% and 67% respectively). Non-sensitivity was defined in accordance to pre-2008 Clinical and Laboratory Standards Institute breakpoints [36]. For both datasets, 'resistance' as used throughout the paper refers to non-sensitivity. The four hospital datasets were obtained from Chang et al. [2] (S2 File). All data were analysed anonymously.

## Linkage disequilibrium

If the frequency of resistance to antibiotic $a$ is $p_a$ and the frequency of resistance to antibiotic $b$ is $p_b$, the coefficient of linkage disequilibrium between resistance to antibiotics $a$ and $b$ is $D_{ab} = p_{ab} - p_a p_b$, where $p_{ab}$ is the frequency of resistance to both $a$ and $b$. The normalised coefficient $D'_{ab}$ is given by: $D'_{ab} = \frac{D_{ab}}{min(p_a p_b, (1-p_a)(1-p_b))}$ if $D_{ab} < 0$ and $D'_{ab} = \frac{D_{ab}}{min(p_a(1-p_b), (1-p_a)p_b)}$ if $D_{ab} > 0$.

In general the sign of $D'$ is arbitrary because it depends on which alleles are chosen for the calculation. We consistently calculate $D'$ using the frequency of resistance: positive $D'$ therefore means resistance to one antibiotic is associated with resistance to the other, while negative $D'$ means association between sensitivity and resistance.

## Model implementation and code availability

All described models were implemented in Wolfram Mathematica (version 11.2.0.0). Modelling results are numerical solutions at t = 100000 (equilibrium is reached considerably earlier, see Fig H in S1 Text). For computing $D'$, numerical results for strain frequencies have been rounded to the nearest $10^{-10}$ to ensure strain frequencies for absent strains are zero (as opposed to zero within numerical error). The code is provided as a supporting file.

## Effect of intergroup transmission and recombination

To test the effect of relaxing the assumption that the pathogen dynamics can be divided into non-interacting sub-models, we include three additional models.

First, we model the dynamics of resistance to three antibiotics (i.e. eight possible resistance profiles) spreading in a host population consisting of five host groups. The antibiotics make up different proportions of total antibiotic consumption (20, 35 and 45% of total antibiotic consumption rate $\tau$). The pathogen experiences a different clearance rate within each host class $p$ ($\mu_p$). In addition, sub-strain with resistance profile $g$ experiences clearance from antibiotic exposure at rate $\tau_g$ which depends on its resistance status: $\tau_g = \tau(i_a 0.20 + i_b 0.35 + i_c 0.45)$, where $i_a = 1$ if $g$ is sensitive to antibiotic $a$ and 0 otherwise. Resistance to each antibiotic decreases transmission rate by a factor of $c$. Uninfected hosts of class $p$ ($U_p$) are therefore infected at rate $c^{n_g}\beta\left[(1-m)I_{g,p} + \frac{m}{4}\sum_{x\in P'} I_{g,x}\right]$, where $n_g$ is the number of antibiotics strain $g$ is resistant to, $m$ is a parameter that sets the extent of mixing between the classes and $P'$ is the set of population classes excluding $p$. The dynamics of strain $g$ within population $p$ are thus described by:

$$\frac{dI_{g,p}}{dt} = c^{n_g}\beta\left[(1-m)I_{g,p} + \frac{m}{4}\sum_{x\in P'} I_{g,x}\right]U_p - (\tau_g + \mu_p)I_{g,p} \quad (10)$$

Second, we model the dynamics of resistance to three antibiotics in a single host population in pathogen with five strains differing in clearance rate (i.e. eight resistance profiles and five strains, giving a total of 40 possible sub-strains) with recombination at the duration of carriage

locus. Strain $i$ is cleared at rate $\mu_i$ and, as above, sub-strains with resistance profile $g$ experience clearance from antibiotic exposure at rate $\tau_g$ which depends on its resistance status: $\tau_g = \tau(i_a 0.20 + i_b 0.35 + i_c 0.45)$. Resistance to each antibiotic decreases transmission rate by a factor of $c$. Balancing selection is modelled similarly to Lehtinen et al. [20], by scaling transmission rate of strain $i$ by a factor $\psi_i$ which depends on the strain's prevalence: $\psi_i = (1 - [\frac{\sum_x I_{x,i}}{1-U} - \frac{1}{5}])^k$, where $k$ is a parameter setting the strength of balancing selection and $U$ is the uninfected host class. Recombination at the duration of carriage locus is modelled by allowing hosts infected with strain $i$ with resistance profile $g$ to transmit strain $j$ with resistance profile $g$ at a rate $r\Sigma_x I_{x,j}$. Recombination therefore decreases the transmission of strain $i$ with resistance profile $g$ by $\rho_{g,i} = rI_{g,i} \Sigma_x \Sigma_y I_{x,y}$ and increases it by $\kappa_{g,i} = r\Sigma_y \Sigma_x I_{g,y} I_{x,i}$. Note that the recombination rate parameter $r$ captures the probability of co-infection, the probability of recombination occurring and the probability of transmitting the recombinant sub-strain. The dynamics of strain $i$ with resistance profile $g$ are described by:

$$\frac{dI_{g,i}}{dt} = c^{n_g}\psi_i\beta\left[I_{g,i} - \rho_{g,i} + \kappa_{g,i}\right]U - (\tau_g + \mu_i)I_{g,i} \tag{11}$$

The third model is the same as the one above, with the exception that recombination occurs at the resistance loci instead of the duration of carriage locus. It is therefore described by Eq (11), but the expressions for $\rho$ and $\kappa$ are different. We define resistance profile $g'_a$ as a resistance profile otherwise identical to $g$, but with the other allele at locus $a$ (i.e. if $g$ is sensitive to antibiotic $a$, $g'_a$ is resistant), $N_{g,a}$ as the set of resistance profiles with the same allele at locus $a$ as profile $g$ and $N'_{g,a}$ as the set of resistance profiles with the different allele at locus $a$ than profile $g$. Hosts infected with strain $i$ with resistance profile $g$ transmit a strain $i$ with a resistance profile $g'_a$ at rate $r\sum_j \sum_{x \in N'_{g,a}} I_{x,j}$. Recombination can occur at any of the three resistance loci (we assume recombination rates are low enough to ignore the possibility of recombination occurring at multiple loci at the same time). Recombination therefore decreases the transmission of strain $i$ with resistance profile $g$ by $\rho_{g,i} = 3rI_{g,i} \Sigma_x \Sigma_y I_{x,y}$ and increases it by $\kappa_{g,i} = r(I_{g'_a,i}\sum_y \sum_{x \in N_{g,a}} I_{x,y} + I_{g'_b,i}\sum_y \sum_{x \in N_{g,b}} I_{x,y} + I_{g'_c,i}\sum_y \sum_{x \in N_{g,c}} I_{x,y})$.

The parameter values for the results presented in Fig 3 are: $c = 0.95$, $\beta = 2$, $\{\mu_1, .., \mu_5\} = \{1.2, 1., 0.8, 0.6, 0.4\}$, $\tau = 0.12$ and $k = 5$.

## Effect of imperfectly correlated antibiotic consumption rates

To test the effect of relaxing the assumption that all host groups consume different types of antibiotics in identical proportions, we model the dynamics of resistance to two antibiotics in a population consisting of ten host groups. The dynamics within each host group are represented by Eq 3, with parameter values $c_\beta = 0.95$ and $c_\mu = 1$ for both antibiotics, $\beta = 2$, and, unless otherwise stated $\mu = 1$. There is no transmission between these host groups. For both drugs, five of these host groups consume the antibiotic at a rate which selects for resistance when $\mu = 1$ ($\tau_{high} = 0.075$), and five at a rate which selects for sensitivity when $\mu = 1$ ($\tau_{low} = 0.025$). There are therefore six different ways in which the consumption rates of the two antibiotics can be combined: all populations consuming the first drug at a high rate also consume the second drug at high rate (Spearman's rho: 1); four out of the five populations consuming the first drug at a high rate consume the second drug at a high rate (Spearman's rho: 0.6); etc. We run a simulation for each of these six possible configurations. To test the effect of additional variation in the resistance proneness of strata, we introduce variation in the clearance rate of these populations: the five host groups consuming the first drug at the high rate now have different clearance rates (evenly spaced between a maximum and minimum clearance rate), and similarly for the five

host groups consuming the first drug at the low rate. We run a simulation for each of these possible ways the consumption of the second drug can be distributed among these host groups.

## Supporting information

**S1 Text. Supporting text and figures.** PDF file containing all supporting text and figures.
(PDF)

**S1 File. Resistance profiles and duration of carriage estimates for the Maela dataset.** Obtained from [28].
(CSV)

**S2 File. Resistance profiles for the four hospital datasets.** Obtained from [2].
(XLSX)

**S3 File. Code for model simulations.**
(NB)

## Acknowledgments

Members of the Maela Pneumococcal Collaboration: Stephen D. Bentley (Wellcome Sanger Institute, Hinxton, UK), Nicholas J. Croucher (Imperial College London, London, UK), John A. Lees (Wellcome Sanger Institute, Hinxton, UK; NYU School of Medicine, New York, New York, USA), Paul Turner (Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK).

## Author Contributions

**Conceptualization:** Sonja Lehtinen, Christophe Fraser.

**Investigation:** Sonja Lehtinen, François Blanquart, Marc Lipsitch, Christophe Fraser.

**Writing – original draft:** Sonja Lehtinen.

**Writing – review & editing:** Sonja Lehtinen, François Blanquart, Marc Lipsitch, Christophe Fraser.

## References

1. Giske CG, Monnet DL, Cars O, Carmeli Y, ReAct-Action on Antibiotic Resistance. Clinical and Economic Impact of Common Multidrug-Resistant Gram-Negative Bacilli. Antimicrobial Agents and Chemotherapy. 2008; 52(3):813–821. https://doi.org/10.1128/AAC.01169-07 PMID: 18070961
2. Chang HH, Cohen T, Grad YH, Hanage WP, O'Brien TF, Lipsitch M. Origin and proliferation of multiple-drug resistance in bacterial pathogens. Microbiology and Molecular Biology Reviews: MMBR. 2015; 79 (1):101–16. https://doi.org/10.1128/MMBR.00039-14 PMID: 25652543
3. Falagas ME, Bliziotis IA. Pandrug-resistant Gram-negative bacteria: the dawn of the post-antibiotic era? International Journal of Antimicrobial Agents. 2007; 29(6):630–636. https://doi.org/10.1016/j.ijantimicag.2006.12.012 PMID: 17306965
4. Poole K. Efflux-mediated antimicrobial resistance. Journal of Antimicrobial Chemotherapy. 2005; 56 (1):20–51. https://doi.org/10.1093/jac/dki171 PMID: 15914491
5. Piddock LJV. Clinically relevant chromosomally encoded multidrug resistance efflux pumps in bacteria. Clinical Microbiology Reviews. 2006; 19(2):382–402. https://doi.org/10.1128/CMR.19.2.382-402.2006 PMID: 16614254
6. Slatkin M. Linkage disequilibrium?understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics. 2008; 9(6):477. https://doi.org/10.1038/nrg2361 PMID: 18427557

7. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid Pneumococcal Evolution in Response to Clinical Interventions. Science. 2011; 331(6016):430–434. https://doi.org/10.1126/science.1198545 PMID: 21273480
8. Enne VI, Livermore DM, Stephens P, Hall LM. Persistence of sulphonamide resistance in Escherichia coli in the UK despite national prescribing restriction. The Lancet. 2001; 357(9265):1325–1328. https://doi.org/10.1016/S0140-6736(00)04519-0
9. Bean DC, Livermore DM, Papa I, Hall LM. Resistance among Escherichia coli to sulphonamides and other antimicrobials now little used in man. Journal of Antimicrobial Chemotherapy. 2005; 56(5):962–964. https://doi.org/10.1093/jac/dki332 PMID: 16150859
10. Currie CJ, Berni E, Jenkins-Jones S, Poole CD, Ouwens M, Driessen S, et al. Antibiotic treatment failure in four common infections in UK primary care 1991-2012: longitudinal analysis. BMJ. 2014; 349: g5493. https://doi.org/10.1136/bmj.g5493 PMID: 25249162
11. Dolk FCK, Pouwels KB, Smith DR, Robotham JV, Smieszek T. Antibiotics in primary care in England: which antibiotics are prescribed and for which conditions? Journal of Antimicrobial Chemotherapy. 2018; 73(suppl_2):ii2–ii10. https://doi.org/10.1093/jac/dkx504 PMID: 29490062
12. Shallcross L, Beckley N, Rait G, Hayward A, Petersen I. Antibiotic prescribing frequency amongst patients in primary care: a cohort study using electronic health records. Journal of Antimicrobial Chemotherapy. 2017; 72(6):1818–1824. https://doi.org/10.1093/jac/dkx048 PMID: 28333200
13. Tedijanto C, Olesen SW, Grad YH, Lipsitch M. Estimating the proportion of bystander selection for antibiotic resistance among potentially pathogenic bacterial flora. Proceedings of the National Academy of Sciences. 2018; 115(51):E11988–E11995. https://doi.org/10.1073/pnas.1810840115
14. Ward H, Perron G, Maclean R. The cost of multiple drug resistance in Pseudomonas aeruginosa. Journal of Evolutionary Biology. 2009; 22(5):997–1003. https://doi.org/10.1111/j.1420-9101.2009.01712.x PMID: 19298493
15. Angst DC, Hall AR. The cost of antibiotic resistance depends on evolutionary history in Escherichia coli. BMC Evolutionary Biology. 2013; 13(1):163. https://doi.org/10.1186/1471-2148-13-163 PMID: 23914906
16. Trindade S, Sousa A, Xavier KB, Dionisio F, Ferreira MG, Gordo I. Positive Epistasis Drives the Acquisition of Multidrug Resistance. PLoS Genetics. 2009; 5(7):e1000578. https://doi.org/10.1371/journal.pgen.1000578 PMID: 19629166
17. Borrell S, Teo Y, Giardina F, Streicher EM, Klopper M, Feldmann J, et al. Epistasis between antibiotic resistance mutations drives the evolution of extensively drug-resistant tuberculosis. Evolution, Medicine, and Public Health. 2013; 2013(1):65–74. https://doi.org/10.1093/emph/eot003 PMID: 24481187
18. Wong A. Epistasis and the Evolution of Antimicrobial Resistance. Frontiers in Microbiology. 2017; 8:246. https://doi.org/10.3389/fmicb.2017.00246 PMID: 28261193
19. Colijn C, Cohen T, Fraser C, Hanage W, Goldstein E, Givon-Lavi N, et al. What is the mechanism for persistent coexistence of drug-susceptible and drug-resistant strains of Streptococcus pneumoniae? Journal of The Royal Society Interface. 2010; 7(47). https://doi.org/10.1098/rsif.2009.0400
20. Lehtinen S, Blanquart F, Croucher NJ, Turner P, Lipsitch M, Fraser C. Evolution of antibiotic resistance is linked to any genetic mechanism affecting bacterial duration of carriage. Proceedings of the National Academy of Sciences of the United States of America. 2017; 114(5):1075–1080. https://doi.org/10.1073/pnas.1617849114 PMID: 28096340
21. Cobey S, Baskerville EB, Colijn C, Hanage W, Fraser C, Lipsitch M. Host population structure and treatment frequency maintain balancing selection on drug resistance. Journal of The Royal Society Interface. 2017; 14(133). https://doi.org/10.1098/rsif.2017.0295
22. Kouyos R, Klein E, Grenfell B, Degrave M, Vanjak D. Hospital-Community Interactions Foster Coexistence between Methicillin-Resistant Strains of Staphylococcus aureus. PLoS Pathogens. 2013; 9(2): e1003134. https://doi.org/10.1371/journal.ppat.1003134 PMID: 23468619
23. Blanquart F, Lehtinen S, Lipsitch M, Fraser C. The evolution of antibiotic resistance in a structured host population. Journal of The Royal Society Interface. 2018; 15(143):20180040. https://doi.org/10.1098/rsif.2018.0040
24. Davies NG, Flasche S, Jit M, Atkins KE. Within-host dynamics shape antibiotic resistance in commensal bacteria. Nature Ecology & Evolution. 2019; 3(3):440. https://doi.org/10.1038/s41559-018-0786-x
25. Cobey S, Lipsitch M. Niche and neutral effects of acquired immunity permit coexistence of pneumococcal serotypes. Science. 2012; 335(6074):1376–1380. https://doi.org/10.1126/science.1215947 PMID: 22383809
26. Melnyk AH, Wong A, Kassen R. The fitness costs of antibiotic resistance mutations. Evolutionary Applications. 2015; 8(3):273–283. https://doi.org/10.1111/eva.12196 PMID: 25861385

27. Leibovitz E. The use of fluoroquinolones in children. Current Opinion in Pediatrics. 2006; 18(1):64–70. PMID: 16470165

28. Lees JA, Croucher NJ, Goldblatt D, Nosten F, Parkhill J, Turner C, et al. Genome-wide identification of lineage and locus specific variation associated with pneumococcal carriage duration. eLife. 2017; p. e26255. https://doi.org/10.7554/eLife.26255 PMID: 28742023

29. Weinberger DM, Trzciński K, Lu YJ, Bogaert D, Brandes A, Galagan J, et al. Pneumococcal capsular polysaccharide structure predicts serotype prevalence. PLoS Pathogens. 2009; 5(6):e1000476. https://doi.org/10.1371/journal.ppat.1000476 PMID: 19521509

30. Corander J, Fraser C, Gutmann MU, Arnold B, Hanage WP, Bentley SD, et al. Frequency-dependent selection in vaccine-associated pneumococcal population dynamics. Nature Ecology & Evolution. 2017.

31. Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. The dynamics of molecular evolution over 60,000 generations. Nature. 2017; 551:45. https://doi.org/10.1038/nature24287 PMID: 29045390

32. Goldstein E, Kirkcaldy RD, Reshef D, Berman S, Weinstock H, Sabeti P, et al. Factors related to increasing prevalence of resistance to ciprofloxacin and other antimicrobial drugs in Neisseria gonorrhoeae, United States. Emerging Infectious Diseases. 2012; 18(8):1290. https://doi.org/10.3201/eid1808.111202 PMID: 22840274

33. Turner P, Turner C, Jankhot A, Helen N, Lee SJ, Day NP, et al. A longitudinal study of Streptococcus pneumoniae carriage in a cohort of infants and their mothers on the Thailand-Myanmar border. PLOS ONE. 2012; 7(5):e38271. https://doi.org/10.1371/journal.pone.0038271 PMID: 22693610

34. Finkelstein JA, Huang SS, Daniel J, Rifas-Shiman SL, Kleinman K, Goldmann D, et al. Antibiotic-resistant Streptococcus pneumoniae in the heptavalent pneumococcal conjugate vaccine era: predictors of carriage in a multicommunity sample. Pediatrics. 2003; 112(4):862–869. https://doi.org/10.1542/peds.112.4.862 PMID: 14523178

35. Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, et al. Population genomics of post-vaccine changes in pneumococcal epidemiology. Nature Genetics. 2013; 45(6):656–663. https://doi.org/10.1038/ng.2625 PMID: 23644493

36. Jorgensen JH, Turnidge JD. Susceptibility test methods: dilution and disk diffusion methods. In: Manual of Clinical Microbiology, Eleventh Edition. American Society of Microbiology; 2015. p. 1253–1273.