Supplementary Text:

Stochastic bacterial population dynamics prevent the emergence of antibiotic resistance within the mutant selection window

Helen K. Alexander* and R. Craig MacLean
Department of Zoology, University of Oxford, Oxford, UK, OX1 3PS
* helen.alexander@zoo.ox.ac.uk

Part I

Detailed Experimental Methods and Data Processing

1 Distribution of colony-forming units

The goal of this experiment was to test whether the number of viable cells, quantified by CFUs, in a small volume of highly diluted culture can be described by a Poisson distribution. An overnight culture was diluted 10^7-fold in PBS and plated in a 6×8 array of 4μl spots on each of three square (12cm × 12cm) LB-agar plates, yielding a total of 144 replicate spots. These LB-agar plates were incubated for the remainder of the day of plating at 37°C, moved to room temperature overnight, then returned to 37°C in the morning until colonies were visible but still separated for counting.

2 Competition experiment

The goal of this experiment was to determine the direction of selection (for resistance vs. sensitivity) across a range of streptomycin concentrations, by competing resistant and sensitive strains at reasonably high starting densities (such that demographic stochasticity is negligible).

Experimental protocol: We used the YFP-labelled resistant strain (as usual) and the DsRed-labelled sensitive strain. Although DsRed does not provide a strong enough fluorescent signal to aid in discrimination, it ensures that any fitness effect of carrying a fluorescent marker is controlled across

the strains.

Overnight cultures of each strain were mixed at an initial 20-fold dilution each, then this mixture was diluted 100-fold further. This yielded an approximately 1:1 mixture of strains each at 2000-fold dilution, which was inoculated at $20\mu l$ per $200\mu l$ total culture volume. The total bacterial density at the start of treatment is thus expected to be around 5×10^5 CFU/ml, as we also used for our MIC tests at standard inoculation density (see Section 4). For the pure cultures, each strain was diluted 2000-fold alone and then inoculated similarly. That is, we chose to match the density of a given strain between pure and mixed cultures, rather than the total bacterial density.

At each streptomycin concentration, we inoculated a total of 6 replicate mixed cultures and 2 replicates of each pure culture, split evenly across two treatment plates (arbitrarily labelled A and B). Test concentrations ranged from 1/16 to 16 $\times$ MIC_S streptomycin in 2-fold steps, as well as streptomycin-free. Outer wells were also filled with streptomycin-free media and mock-inoculated with 20μl of PBS to buffer against evaporation and to serve as media-only controls. Treatment plates were incubated (37°C, 225rpm), sampled (20μl per well) at 6.5h, then immediately returned to the incubator and sampled again at approx. 24h. The latter time point provided better resolution at higher streptomycin concentrations and is thus used for data analysis. The 24h treatment culture samples (along with media-only controls) were diluted a total of 500-fold in sterile filtered PBS for flow cytometry (BD Accuri C6 Flow Cytometer with software version 1.0.264.21 – Accuri Cytometers, Inc.). We sampled 66μl per well with fast fluidics, i.e. 66μl/min, discarding events with forward scatter FSC-H $< 10,000$ or side scatter SSC-H < 8000.

Data processing: To analyze the flow cytometry data, we proceeded as follows.

1. We defined non-overlapping gates 'S' and 'R' in the FL1 (fluorescence detection) – FSC (forward scatter) plots (see Suppl. Fig. 2). FL1 is configured with a blue (488nm) laser and 533/30 interference filter, which primarily detects the YFP signal. The S and R gates thus roughly correspond to DsRed-labelled sensitive cells (lower fluorescence) and YFP-labelled resistant cells (higher fluorescence), respectively. However, the pure cultures revealed overlap into the opposite gates, particularly of resistant cells with low fluorescence into the 'S' gate, which is accounted for below. The gates were drawn separately, but similarly, for each treatment plate. In total, these gates comprised an average of 98% of all detected events on treatment plate A (range across wells: 95-99%) and an average of 96% (range: 89-98%) on treatment plate B. Events falling outside both gates were excluded from analysis.

2. We corrected for background events in each well by subtracting the number of events in the corresponding media-only control from the number of events in the sample of interest, in each gate. If negative, we set this value to zero.

3. From the (background-adjusted) number of events in each gate in pure cultures, where we know only a single strain is present, we calculated the parameters $p_{i,j}$: the proportion of cells of strain i that fall into gate j. For example, $p_{S,R}$ is the proportion of sensitive cells that fall into the "resistant" gate, calculated as:

$$p_{S,R} = \frac{G_R^{\text{pure } S}}{G_S^{\text{pure } S} + G_R^{\text{pure } S}}$$

where $G_j^{\text{pure } i}$ is the number of events falling into gate j in the pure culture of strain i. (Thus, $p_{S,S} + p_{S,R} = 1$ from the pure sensitive culture and $p_{R,S} + p_{S,S} = 1$ from the pure resistant culture.) These parameters are calculated separately at each streptomycin concentration, but crucially, we assume below that they are the same for a given strain when it is in a mixed culture as in a pure culture.

4. In mixed cultures, we want to know the true number of cells of each strain ($N_{\text{S,tot}}^{\text{mix}}$ and $N_{\text{R,tot}}^{\text{mix}}$), adding up cells that fall into either gate: that is, for each strain i, $N_{\text{i,tot}}^{\text{mix}} = N_{\text{i,S}}^{\text{mix}} + N_{\text{i,R}}^{\text{mix}}$. On the other hand, what we observe in a mixed culture is the total number of cells of either strain that fall into each gate j, G_j^{mix}. We can express the relationship between these quantities as:

$$G_S^{\text{mix}} = p_{S,S} N_{\text{S,tot}}^{\text{mix}} + p_{R,S} N_{\text{R,tot}}^{\text{mix}}$$
$$G_R^{\text{mix}} = p_{S,R} N_{\text{S,tot}}^{\text{mix}} + p_{R,R} N_{\text{R,tot}}^{\text{mix}}$$

where the parameters $p_{i,j}$ were calculated above from the pure cultures. Thus we have two linear equations in two unknowns, which can be readily solved. We obtain:

$$N_{\text{S,tot}}^{(\text{mix})} = \frac{p_{R,R} G_S^{\text{mix}} - p_{R,S} G_R^{\text{mix}}}{p_{S,S} p_{R,R} - p_{S,R} p_{R,S}}$$
$$N_{\text{R,tot}}^{(\text{mix})} = G_S^{\text{mix}} + G_R^{\text{mix}} - N_{\text{S,tot}}^{(\text{mix})} \tag{S1}$$

5. Equation S1 was generally applied to infer the number of cells of each strain in each mixed culture after treatment, with the proportion resistant then calculated as $N_{\text{R,tot}}^{(\text{mix})} / \left(N_{\text{S,tot}}^{(\text{mix})} + N_{\text{R,tot}}^{(\text{mix})} \right)$. A few special cases required adjustments. At the highest streptomycin concentrations, we observed no events above background counts in the pure sensitive-strain cultures; thus the parameters $p_{S,j}$ were undefined. Here we assumed that all events in the mixed cultures at these concentrations were resistant cells. At intermediate streptomycin concentrations, where sensitive cells were at low density but not eradicated, the general formula sometimes returned a small negative value for the number of sensitive cells. This error reflects the imperfect assumption that the proportion of cells falling in each gate is the same across cultures, whereas it will in reality show some variation. In these cases, we manually set the number of sensitive cells to zero; this adjustment had a very small effect relative to the total number of cells.

3 Seeding experiments

Experimental protocol: Media (LB) containing various concentrations of streptomycin was dispensed onto 96-well plates ($180\mu l$/well) using an automated liquid handler (BioTek Precision XS). An overnight culture of the YFP-labelled resistant strain was serially diluted in PBS up to 10^6-fold, then in one further independent step in each case, diluted to 4×10^7-fold, 8×10^7-fold, and 1.6×10^8-fold. These diluted cultures were used to inoculate the treatment plates at $20\mu l$ per well. In parallel, control plates containing streptomycin-free media were mock-inoculated with $20\mu l$/well of PBS to check for contamination, since the automated liquid handler is not maintained in a sterile environment. All plates were incubated (37°C, 225rpm) for 3d, and removed daily to measure optical density (OD_{595}) on a BioTek Synergy 2 plate reader. Lids were removed briefly for this reading; again, control plates served as checks for the extent of contamination.

In this experiment, each 96-well treatment plate corresponded to 96 replicates at the same streptomycin concentration and dilution factor; thus, it is possible that plate effects were confounded with the treatment conditions. This possible source of error is mitigated by pooling data across three dilution factors (three separate plates) at each streptomycin concentration in the likelihood-based model fitting (see Sections 9 and 12). Furthermore, since we obtained similar results in independent experiments (Suppl. Table 1), we can be fairly confident in the effect of streptomycin as opposed to results being dominated by random plate effects.

We note that such a seeding experiment has also been used previously, on a much smaller scale, by Levin-Reisman et al.[4] (see their Supplementary Material). However, those authors did not infer a single-cell establishment probability from the proportion of populations showing growth.

Data processing: As described in the main text Methods, wells were scored as established if their optical density exceeded 0.1. The number of wells eventually (by Day 3) scored as established on each plate was taken as data for model fitting (see Part II).

With streptomycin concentrations up to $1/32\times\text{MIC}_R$, there were no new appearances of growth after Day 1, except in a single well in each experiment on Day 2. At $1/16\times\text{MIC}_R$, growth was first detected on either Day 1 or 2, but there were no further appearances on Day 3. Thus, at these concentrations, we are fairly confident that we allowed sufficient time for growth to allow detection of all established populations. At $1/8\times\text{MIC}_R$, there were still some new appearances on Day 3; thus, it is possible that we missed rare slow-growing wells that would have crossed the OD threshold even later. However, our estimates of establishment probability at $1/16\times\text{MIC}_R$ in the seeding experiments were similar to, and even slightly higher than, our estimates from the later inoculum size effect tests (Section 5) where we read OD up to 5 days post-inoculation (see Suppl. Table 1). Thus, we do not expect that potential undetected growth in the seeding experiment substantially affected our results.

In experiment 2, a partially clogged dispenser needle on the automated liquid handler resulted in a reduced volume of media in row H of each plate (estimated $\sim 150\mu l$ instead of $180\mu l$) and a few of these wells dried up by Day 3. In one case, this resulted in OD that was above the threshold on Day 2 but dropped below the threshold on Day 3. This well was checked visually to confirm bacterial growth and scored as established.

In each of the two experiments, a single well out of two 96-well control plates reached OD > 0.1, first appearing on Day 2 or 3. This indicates a negligible contamination rate of $\sim 0.5\%$.

4 MIC tests

For the MIC tests at variable inoculation density, we used an overnight culture diluted 10^3-, 10^4-, 10^5-, and 10^6-fold to inoculate test cultures at a further 1 in 10 dilution. The lowest dilution factor yielded thus final inoculation density around 5×10^5 CFU/ml, as used in a standard MIC test [1]. Culture growth was evaluated at approximately 20h, 2d and 3d, by $OD_{595} > 0.1$, as described above for the seeding experiments.

Actual inoculum sizes were estimated by plating out diluted overnight culture on round LB-agar plates ($20\mu l$ diluted culture per plate, spread with glass beads) and counting colony-forming units after overnight incubation at 37°C. We averaged the results of 5-6 plates at each of three 5×10^6-fold dilutions of the overnight culture, each taken in one further dilution step from the 10^6-fold dilution used for the smallest inoculum size on the test plates.

In the main experiment (Fig. 4 and Suppl. Fig. 3), we used standard $200\mu l$ cultures on 96-well test plates. In the supplementary experiment, where we co-varied absolute inoculum size and density (Suppl. Fig. 4), we additionally tested $1160\mu l$ cultures on 24-well test plates (Falcon flat-bottom non-treated 24-well cell culture plate with lid, product no. 351147), and $116\mu l$ cultures on 96-well test plates. The $1160\mu l$ volume was chosen to match surface area to volume ratio on the standard test plates. The $116\mu l$ volume was then chosen to obtain 10-fold lower volume and hence 10-fold higher density at matched absolute inoculum sizes.

In both experiments, we tested streptomycin-free positive controls along with streptomycin-containing media in 2-fold concentration steps from $1/16\times\text{MIC}_R$ to $2\times\text{MIC}_R$ (standard volume experiment) or up to $1\times\text{MIC}_R$ (varying volumes experiment, in which the capacity was limited by the 24-well plates). Each experiment scored growth in 6 independent cultures per test condition.

In the standard volume experiment, the four inoculation densities were distributed across four test plates to control for any plate effects. Outer rows on each test plate filled with streptomycin-free media and mock-inoculated with PBS served as negative controls; no contamination was detected over the course of the experiment. In the varying volume experiment, on 96-well test plates, there were two

replicates at each inoculation density along with negative controls on each of three plates; again, no contamination was detected over the course of the experiment. On 24-well test plates, one replicate per inoculation density was tested on each of six plates, while an additional plate processed in parallel served as a negative control and showed no contamination.

Occasionally, a replicate showed no growth at a given streptomycin concentration, but did grow at the next highest concentration step, before growth was abolished. (This occurred in the experiment presented in Fig. 4 and Suppl. Fig. 3 for one replicate at estimated inoculum size 1.28e3 CFU evaluated at 20h, and one replicate at 1.28e4 CFU evaluated at 3d.) This effect can arise through stochastic effects leading to growth in some cultures but not others at any given streptomycin concentration. Furthermore, cultures at successive streptomycin concentrations grew independently of one another, and were grouped simply by plate position as one replicate for MIC evaluation. In these ambiguous cases, we score MIC for the replicate as the higher concentration, at and beyond which no growth occurred.

5 Testing the null model of the inoculum size effect

In this experiment, we evaluated the probability of population growth at a given streptomycin concentration as a function of inoculum size (main Fig. 5 and Suppl. Fig. 5-6). In each experiment, we tested growth in streptomycin-free media in parallel, in order to estimate inoculum size (see Sections 8-9). We made a single dilution series of an overnight culture to inoculate test plates, but selected a different subset of these diluted cultures to use at different streptomycin concentrations (see Table S1). Thus, any inaccuracy in individual steps of the dilution series is a possible source of discrepancy between streptomycin-free and streptomycin-containing cultures. We did not factor this possible error into our model (rather, assumed that dilution steps were perfect), but we minimized the possibility of compounding errors in the experimental protocol by taking dilutions in parallel rather than in series wherever possible (e.g. 1e6- through 2e7-fold dilutions would each be prepared in a single independent step from a common 2e5-fold dilution). To control for possible plate effects, different inoculum sizes at a given streptomycin condition were distributed across test plates. All test plates included edge wells as negative controls to check for contamination.

Growth of streptomycin-free cultures was scored (by $OD_{595} > 0.1$) after one day, which was determined to be sufficient for stabilization of detectable OD. (In the main and first supplementary experiments, no new appearances of growth occurred after the first day. In the second supplementary experiment, there were two new appearances of growth, but also several contaminated wells on the second day; thus, we did not count these new appearances.)

Growth of streptomycin-treated cultures was scored daily up to five days post-inoculation. This

Table S1: Testing the null model of inoculum size effect: experimental set-up

Experiment	[Strep] ($\times$MIC$_R$)	Dilution factors inoculated*	# reps. per dil. fac.
main	0	2e9, 5e8, 2e8, 1e8, 5e7	54
	1/16	5e9, 2e9, 1e9, 5e8, 2e8, 1e8, 5e7, 2e7, 1e7	54
	1/8	5e8, 1e8, 2e7, 5e6, 1e6, 2e5	54
suppl. 1	0	2e9, 5e8, 2e8, 1e8, 5e7	48†
	1/16	2e9, 5e8, 2e8, 1e8, 5e7, 2e7, 1e7, 5e6, 2e6	54
suppl. 2	0	2e9, 5e8, 2e8, 1e8, 5e7	54
	1/8	5e8, 2e8, 1e8, 5e7, 2e7, 1e7, 5e6, 2e6, 1e6, 2e5	54

* Dilution factors applied to the overnight culture, inoculated at 20μl per 200μl total culture volume.

† Number of replicates reduced because one test plate was dropped early in the experiment.

extended protocol was chosen since previously in the seeding experiments we observed some new appearances of growth on the third day at the highest streptomycin concentration (1/8$\times$MIC$_R$). Although we generally scored growth appearing up to Day 5 here, we observed few new appearances beyond Day 3. Occasionally a well dried up due to evaporation by Day 5; in these cases, we counted culture growth if it appeared earlier. In supplementary experiment 2, a cluster of four wells on one plate hovered just below the threshold OD of 0.1, with one of these wells just crossing the threshold on two of the five measurements. However, there did not appear to be bacterial growth in these wells and thus they were not counted.

Contamination was generally rare, and when it did appear, it was usually late in the experiment (likely due to contamination during OD measurements with lids removed) and judged unlikely to have affected our results. There was one possible exception in supplementary experiment 2, in which there were four new appearances of growth in test cultures on Day 5, but also two cases of contamination appearing in adjacent wells on one plate. In this case, we repeated the model fitting on growth data assessed at Day 4 instead of Day 5, and still found an acceptable fit of the null model ($D = 11.5$, $p = 0.24$, cf. Table S2).

6 Fraction of dead cells (live-dead staining and flow cytometry)

The goal of this experiment was to assess the proportion of dead cells induced by sub-MIC$_R$ streptomycin treatment of the resistant strain.

Experimental protocol: We used the resistant strain without any fluorescent label, so as not to interfere with the fluorescent signal from the live-dead stain. Treatment cultures on 96-well plates were inoculated ($20\mu l$ per $200\mu l$ total culture volume) with a 10^3-fold diluted overnight culture, as in the MIC tests at standard density. Media-only controls were mock-inoculated with the same volume of PBS. We tested streptomycin-free, 1/64, 1/32, 1/16, and $1/8 \times \mathrm{MIC}_R$ streptomycin with six replicate cultures per concentration. The treatment plate was incubated (37°C, 225rpm) for 7h, then cultures were immediately diluted 10-fold into sterile filtered PBS. One additional streptomycin-free culture was sampled and heat-killed (10min at 70°C) before likewise diluting 10-fold.

Live-dead staining and flow cytometry were carried out with one set of replicates at a time in order to avoid having samples exposed to the stain for too long. Each of the six replicate sets included a media-only control, a heat-killed control, and a culture treated at each tested streptomycin concentration; the streptomycin-free culture was also repeated as the last sample in order to check for an effect of time exposed to the stain before sampling. For each replicate set, the 10-fold diluted samples were diluted a further 10-fold into pre-warmed, sterile filtered 1mM EDTA in PBS and incubated for 10min at 37°C. Then $2\mu l$ each of thiazole orange [TO] and propidium iodide [PI] (BD Cell Viability Kit, product no. 349483) was added per $200\mu l$ sample and incubated 5min further at room temperature. We then sampled $50\mu l$ per sample using flow cytometry (BD Accuri C6 Flow Cytometer; fast fluidics; discarding events with forward scatter FSC-H $< 10{,}000$ or side scatter SSC-H < 8000).

Data processing: To analyze the flow cytometry data, we proceeded as follows.

1. Cell densities in treated and diluted cultures were sometimes low, especially at higher streptomycin concentrations. In order to better discriminate cells from background, we first defined a gate (labelled "cells") in the FSC/SSC (forward/side scatter) plot that incorporated the majority of events in the sampled cultures, but excluded the majority of events in the media-only controls (see Suppl. Fig. 7a). Further analysis was limited to events within this gate.

2. We then defined gates around events in the FL1/FL3 plot that clustered according to their fluorescence (see Suppl. Fig. 7b). TO (live stain) is primarily detected in the FL1 channel (488nm laser with 533/30 filter), while PI (dead stain) is primarily detected in FL3 (488nm laser with 670LP filter). Thus a cluster appearing with higher FL1 and lower FL3 was labelled "intact" and a cluster appearing with lower FL1 and higher FL3 was labelled "dead". Nearly all events in the heat-killed controls fell within the "dead" gate. Together, these two gates incorporated the majority of events in the sampled cultures, but only a minority of events in the media-only controls, providing further discrimination from background events.

3. Finally, to correct for any remaining background, within each replicate set we subtracted the number of events in the media-only control from the number of events in each sampled culture, within each of the two gates ("dead" and "intact").

4. The proportion of dead cells in sampled cultures was defined as the number of events falling within the "dead" gate divided by the total number falling in either the "dead" or the "intact" gate (after background correction).

TO has toxic effects on cells, and thus the order of sampling the diluted cultures by flow cytometry within each replicate set (hence time exposed to the stain before sampling) could potentially have been confounded with the effect of streptomycin treatment. However, by comparing the streptomycin-free culture sampled earlier vs. later within each replicate set, we found that the proportion of dead cells on average actually decreased (mean of six replicates: 0.040 in first sample vs. 0.026 in second), though this difference is not significant (two-sided paired t-test: $p = 0.09$). Thus, changes in the proportion of dead cells can be attributed to streptomycin treatment rather than duration of staining.

7 Viable cell population dynamics

In this experiment, we tracked the dynamics of viable cells over time by plating out samples of cultures treated with various concentrations of streptomycin.

Experimental protocol: Treatment plates were split into Set A (containing $1/32\times\text{MIC}_R$ and $1/16\times\text{MIC}_R$ streptomycin treatments as well as streptomycin-free controls) and Set B ($1/8\times\text{MIC}_R$ and $1/4\times\text{MIC}_R$ and streptomycin-free controls). Importantly, we inoculated multiple treatment plates in order to sample a separate plate at each time point; thus, all replicates are independent of one another over time.

An overnight culture was diluted 5×10^5-fold and used to inoculate treatment plates ($20\mu l$ per $200\mu l$ total culture volume), with six independent replicate cultures at each streptomycin concentration (twelve in streptomycin-free media) on each plate. Treatment plates were incubated at 37°C, 225rpm. In each treatment set (A/B), one plate was taken for sampling immediately after inoculation to give a time point close to 0h, in order to assess initial population size using the same method as at all later time points. Subsequent target sampling times were chosen differently for Set A (1h, 2h, 2h 20min, 2h 40min, 3h, 3h 30min, 4h) and Set B (hourly from 1h to 8h, and at approx. 24h) to account for slower growth at higher streptomycin concentrations. In total, we thus had 8 treatment plates in Set A and 10 in Set B. Actual sampling times (time elapsed between inoculation and plating) were recorded in the course of the experiment.

Upon sampling, the treated cultures were plated undiluted in $4\mu l$ spots on each of five square (12cm × 12cm) LB-agar plates, for a total sampling volume of $20\mu l$ out of each $200\mu l$ culture. After sampling, treatment plates were returned to the incubator for later OD reading (after approx. 1, 2, and 3 days) to assess eventual growth in all the sampled plates, as in our previous experiments. Contamination was rare (one non-inoculated edge well on each of two plates showed contamination first appearing on Day 2, out of a total of 36 edge wells/plate × 18 plates).

LB-agar plates were immediately moved to 37°C for the rest of the day, then removed to the bench (room temperature) overnight to prevent overgrowth of colonies, then returned to 37°C the following day for several hours until colonies were visible, but still separated, for optimal counting by eye. Total colony counts from the five plates were used to estimate viable population size in the treated cultures at time of sampling (scaling up by a factor 10 from sampled to total volume). Later plated time points were excluded if colonies became too dense to count at a given streptomycin concentration.

Data processing: For the purpose of statistically testing the effects of streptomycin, time, and their interaction on population size (ANOVA and post-hoc Dunnett's test), we counted sampling times from Sets A and B that were within approximately 10min of each other as the same categorical sampled time. However, precise sampling times were used for plotting in Fig. 6b.

Part II

Mathematical modelling and model fitting

Here we describe the models that we fit to population growth data in the seeding experiments and the tests of the inoculum size effect. We note that our modelling approach and fitting methods are not entirely novel, but present them here in full for clarity. Connections to previous work are briefly discussed at the end (Section 8.2).

8 Theoretical model of population growth

We treat the number of populations showing growth, across independent biological replicates in a given test condition, as binomially distributed with number of trials equal to the number of replicate cultures and 'success' probability p_w, the probability of population growth in a replicate culture. The parameter p_w will depend on the inoculum size, whose expected value $\bar{N}$ is controlled by varying dilution factors of the inoculating culture; and the environmental (media) conditions on the test plates, x. In our case, environment x represents the concentration of streptomycin, but the model and methods are equally applicable to any other environmental factor(s) being tested.

Our fundamental assumption is that population growth will be observed if and only if at least one individual in the inoculum establishes a surviving lineage. (Here we equate 'individual' with 'cell', but generally an individual could be a clump of cells; see Section 8.1 below.) Furthermore assuming that the number of individuals that establish is Poisson distributed with mean α, we can express our model in its most general form as:

$$p_w(\bar{N}, x) = 1 - \mathrm{e}^{-\alpha(\bar{N},x)} \tag{S2}$$

(corresponding to Eqn. 1 in the main text). In support of this assumption, on solid media (streptomycin-free LB-agar), we find that the number of colony-forming units is indeed well described by a Poisson distribution (Suppl. Fig. 1). Note that Equation S2 defines a one-to-one mapping between p_w and α, and thus maximum likelihood estimates and boundaries of confidence intervals on p_w can be transformed to the corresponding results on α.

The "full model" (statistically speaking) involves estimating a separate parameter p_w, or equivalently α, for each $(\bar{N}, x)$ condition tested in the experiment. In this case, the maximum likelihood estimate of p_w is simply the proportion of replicates that show growth (a standard results for the binomial model). The number of parameters in the full model is equal to the number of test conditions, i.e. $|\{(\bar{N}, x)\}|$. We refer to this full model in the original (α) parameterization as **Model A** (see Figure S1).

It is useful to define the relative establishment probability, $\tilde{p}_c$, in a focal environment x (at mean inoculum size $\bar{N}_i$), as the mean number of established cells in that environment, normalized by the result in some baseline environment (for our purposes, streptomycin-free media), defined as $x = 0$:

$$\tilde{p}_c^{(i)}(x) := \frac{\alpha(\bar{N}_i, x)}{\alpha(\bar{N}_i, 0)} \tag{S3}$$

Note that it is possible for $\tilde{p}_c$ to exceed one, if α in environment x exceeds that in environment 0; thus $\tilde{p}_c$ is not a true probability. In practice, estimating $\{\tilde{p}_c^{(i)}(x)\}$ in the full model requires that we have tested growth in both baseline and focal environments at the same mean inoculum size $\bar{N}_i$. (This is the case in the "seeding" experiments, but not in the tests of the inoculum size effect.) We can then rewrite the full model in terms of the transformed parameters $\{\alpha(\bar{N}_i, 0), \{\tilde{p}_c^{(i)}(x_j)\}_{j=1}^{s}\}_{i=1}^{m}$, where $m = |\{\bar{N}\}|$ and $1 + s = |\{x\}|$. Note that the total number of parameters remains the same, with a one-to-one mapping between the original and transformed parameterizations. We call this rewritten version of the full model **Model A′**.

Using these transformed parameters, we can also make the reasonable simplifying assumption that only $\alpha(\bar{N}_i, 0)$ varies with $\bar{N}_i$, while the relative establishment probability $\tilde{p}_c(x_j)$ is constant in a given test environment x_j (**Model B′**). That is, we jointly estimate $\{\alpha(\bar{N}_i, 0)\}_{i=1}^{m}$ and $\{\tilde{p}_c(x_j)\}_{j=1}^{s}$ from the results pooled across all test conditions. Note that this model still requires us to have tested growth in

every environment at the same inoculum sizes. The number of parameters to be estimated is reduced from $m \cdot (s+1)$ in Model A$'$, to $m+s$ in the nested Model B$'$. Similarly, one could pool data across multiple experiments using such a framework, by supposing that $\alpha(\bar{N}_i, 0)$ varies by experiment, but $\tilde{p}_c(x)$ in a given environment remains the same.

Finally, we introduce our null model of the inoculum size effect, i.e. the relationship between p_w and $\bar{N}$. Here we invoke the key assumption that each individual acts independently, i.e. the outcome of establishing a surviving lineage is not affected by other individuals in the inoculum. (This independence assumption is very common when modelling dynamics at low population density, for instance using branching processes; here we will rigorously test the validity of this assumption.) Suppose that the number of individuals in the inoculum is Poisson-distributed with mean $\bar{N}$ and the fate of each individual is an independent Bernoulli trial with success probability p_c ("per-cell establishment probability"), which depends only on the environment, x. Then we arrive at the number of established lineages being Poisson-distributed,[1] consistent with our earlier assumption, and can write the mean very simply as:

$$\alpha(\bar{N}, x) = \bar{N} \cdot p_c(x) \tag{S4}$$

This leads to Eqn. 2 in the main text, which we call the null model, relating p_w to $\bar{N}$. Under this model, inoculum size cancels out in the definition of relative establishment probability (Equation S3) and we have simply:

$$\tilde{p}_c(x) = \frac{p_c(x)}{p_c(0)} \tag{S5}$$

giving the intuitive interpretation that $\tilde{p}_c(x)$ represents the per-cell establishment probability in environment x, normalized by that in the baseline environment.

Note that according to this model, we cannot obtain estimates of absolute establishment probability (p_c) since this parameter plays a symmetrical role to inoculum size ($\bar{N}$) and thus their effects cannot

[1] The number of cells in the inoculum, N, is Poisson-distributed with mean $\bar{N}$ and thus has probability generating function (PGF) $g_N(z) = \mathrm{e}^{-\bar{N}(1-z)}$. Each cell independently has probability p_c of establishing a surviving lineage. Letting Y denote the number of established lineages, $Y|N$ is the sum of N independent Bernoulli(p_c) trials, and thus has a Binomial(N, p_c) distribution with PGF $g_{Y|N}(z) = (1 - p_c + p_c z)^N$. We can then derive the distribution of Y via its PGF, $g_Y(z)$, as follows:

$$\begin{aligned} g_Y(z) := \mathbb{E}\left[z^Y\right] &= \mathbb{E}_N\left[\mathbb{E}\left[z^Y|N\right]\right] \\ &= \mathbb{E}_N\left[(1 - p_c + p_c z)^N\right] \\ &= g_N(1 - p_c + p_c z) \\ &= \mathrm{e}^{-\bar{N}p_c(1-z)} \end{aligned}$$

This is the PGF of a Poisson random variable with mean $\bar{N}p_c$.

be separated. That is, if we observe a higher proportion of established populations in an experiment, we cannot tell whether this was due to higher inoculum size or higher establishment probability. This limitation is not unique to our experimental protocol. Quantifying cell density by counting colony-forming units on agar implicitly assumes that the establishment probability of a "viable cell" is one. Likewise, we will refer to the mean number of established cells in our baseline (growth-optimal) environment, i.e. $\alpha(\bar{N}, 0) = \bar{N}p_c(0)$, as the effective mean inoculum size, and estimate establishment probabilities in all other environments relative to this. The relative establishment probability in environment x, $\tilde{p}_c(x)$, generally provides an upper bound on the absolute establishment probability $p_c(x)$, and if $p_c(0)$ is close to one, as we expect in benign conditions, then $\tilde{p}_c(x)$ will be close to $p_c(x)$.

For the purposes of parameter estimation, we pool results across inoculum sizes by supposing that we do not make any experimental error in culture dilution steps, and so the mean inoculum size is inversely proportional to the dilution factor applied to the inoculating culture.That is, the i^{th} inoculum size is:

$$\bar{N}_i = \bar{N}^*/(d_i/d^*)$$

and thus

$$\alpha(\bar{N}_i, x) = \alpha(\bar{N}^*, x)/(d_i/d^*)$$

where $\bar{N}^*$ is the mean inoculum size at a chosen normalizing dilution factor d^*, and d_i is the i^{th} dilution factor. This assumption can be applied either to the parameterization in terms of $\alpha(\bar{N}^*, x)$ (giving **Model C**) or the transformed parameterization in terms of $\alpha(\bar{N}^*, 0)$ and $\{\tilde{p}_c(x)\}$ (giving **Model C′**), resulting in $|\{x\}|$ parameters to estimate. Since we have now defined a scaling relationship between inoculum sizes, we are no longer constrained to using the same set of inoculum sizes in every environment in order to estimate $\tilde{p}_c$. Note that any deviations of the data from this model fit could reflect not only lack of independence among individuals (the null model assumption), but also experimental errors in the dilution steps.

8.1 Heterogeneous establishment probability

In the above model, we assumed that establishment probability p_c is the same for every individual. However, in reality it may well be the case that cells are in variable physiological states (metabolism, gene expression levels, phases of the cell cycle, etc.) that could affect their division and/or death rates. Furthermore, bacterial cells may aggregate[6], such that the individual units are actually clumps containing variable numbers of cells. Both of this issues imply that the establishment probability should vary among individuals. Here we show mathematically that these issues do not affect our modelling approach and simply require the appropriate interpretation of the parameter p_c.

The key requirement of our model is simply that the number of individual units – regardless of

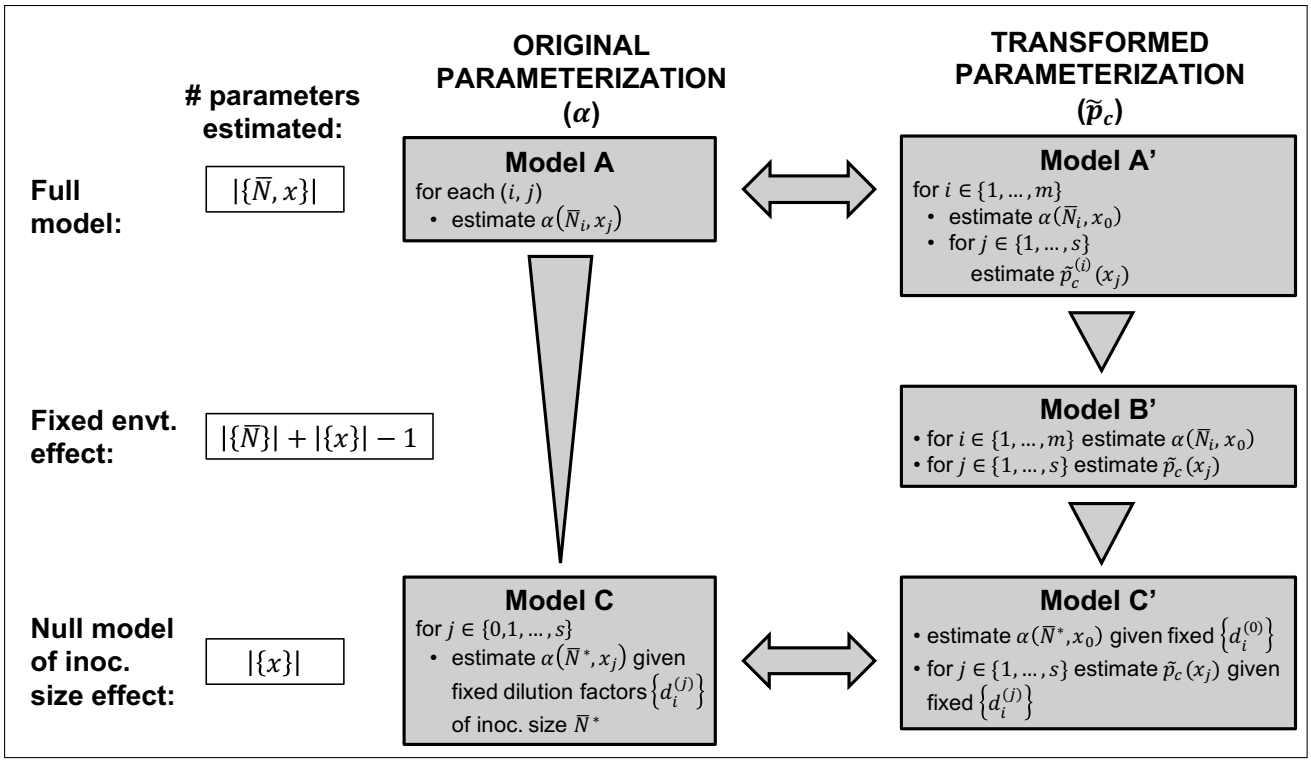

Figure S1: Flow chart summarizing the relationships among models fit to population growth data across multiple inoculum sizes (indexed by i) and environmental conditions (indexed by j). For short, we write m for the number of inoculum sizes tested, i.e. $m = |\{\bar{N}\}|$, and s for the number of non-baseline environments, i.e. $s = |\{x\}| - 1$. Mathematical equivalence of two models is denoted by a double arrow, while a nesting relationship is indicated by a triangle pointing towards the nested model. Note that fitting Model A′ or B′ requires that we have tested every inoculum size × environment combination, while the other models allow different inoculum sizes in different environments.

whether these are single cells or aggregates – is Poisson-distributed. (Recall that we have shown that the number of colony-forming units indeed appears to be Poisson-distributed: Suppl. Fig. 1.) Suppose each of these individuals establishes with probability P, now treated as a random variable drawn independently for each individual from any given distribution with probability density function f_P. For instance, the distribution of P could reflect differences in the phenotypic level of resistance based on intracellular state, or variation in cell aggregate size (with aggregates of more cells likely corresponding to higher establishment probability).

Denoting the number of established lineages derived from an individual as Y (taking on a value of 0 or 1), we have $Y|P \sim \text{Bernoulli}(P)$. The distribution of Y can then be derived as:

$$\begin{aligned}\Pr(Y=1) &= \int f_P(p)\Pr(Y=1|P=p)\,\mathrm{d}p \\ &= \int f_P(p)\,p\,\mathrm{d}p \\ &\equiv \mathbb{E}[P]\end{aligned}$$

Thus, Y is overall a Bernoulli trial with success probability equal to the mean establishment probability, and in turn, the sum of N such independent trials yields a binomial distribution. Therefore, our previous mathematical results still hold if we simply interpret p_c as the mean establishment probability among individual units behaving independently.

8.2 Comparison to previous work

Our theoretical model of population growth is not unique. Indeed, the form of the null model of inoculum size effect,

$$p_w = 1 - \mathrm{e}^{-\bar{N}\cdot p_c}$$

arises very generally under the assumptions that the number of individuals is Poisson-distributed and the fate of each individual lineage is independent[5]. Martin et al.[5] used this general equation to describe the probability of evolutionary rescue in a variety of scenarios (their Equation 3.1). In this case, p_c (or θ_R^* in their notation) represents "rate of rescue per inoculated individual" (ref.[5], p. 4), which captures different processes depending on the scenario. For instance, in the case of rescue relying on *de novo* mutations, θ_R^* accounts for the mutation rate as well as the establishment probability of mutants. In the case of rescue relying of pre-existing (but rare) mutations, θ_R^* is simply the per-individual establishment probability, as in our interpretation. However, we note that this probability could generally differ for a rescue (resistant) variant establishing in isolation, as we tested here, compared to establishment in the presence of an initially large but declining wild-type (sensitive) population, as in the rescue situation.

Importantly, Martin et al.[5] also fit this equation to experimental data using similar methods to ours (cf. Sections 9-10 below). They likewise assess the goodness of fit of this model according to deviance from the full model (our Model C vs. A), but do not include our additional Model B. Our approach moreover differs in that we estimate effective inoculum size from growth in control conditions in parallel with test conditions, rather than treating inoculum size as an independently known value; and we thus derive confidence intervals on the estimated establishment probability relative to the control conditions, that take into account the uncertainty in both measures. As mentioned above, provided establishment is more likely in control than in test conditions, relative establishment probability will remain less than one (up to experimental variation); and provided establishment is very likely in control conditions, relative establishment probability will be close to true establishment probability (which is not separately identifiable in this model).

9 Likelihood-based parameter estimation and model comparison

Basic binomial likelihood: The fundamental model is that the number of populations showing growth at a given test condition, n_{grow}, is binomially distributed with number of trials equal to the total number of populations, n_{tot}, and success probability p_w, the parameter we want to estimate. That is,

$$\Pr(n_{\text{grow}}|(n_{\text{tot}}, p_w)) = \binom{n_{\text{tot}}}{n_{\text{grow}}} p_w^{n_{\text{grow}}} (1-p_w)^{n_{\text{tot}}-n_{\text{grow}}}$$

and so the log likelihood function of p_w given the data $(n_{\text{grow}}, n_{\text{tot}})$ can be written (up to a constant that can be dropped) as:

$$\log \mathcal{L}(p_w|(n_{\text{grow}}, n_{\text{tot}})) = \begin{cases} n_{\text{grow}} \log(p_w) + (n_{\text{tot}} - n_{\text{grow}}) \log(1-p_w)\,, & 0 < n_{\text{grow}} < n_{\text{tot}} \\ n_{\text{tot}} \log(1-p_w)\,, & n_{\text{grow}} = 0 \\ n_{\text{tot}} \log(p_w)\,, & n_{\text{grow}} = n_{\text{tot}} \end{cases} \tag{S6}$$

We have the simple analytical result for the maximum likelihood estimate (MLE):

$$\hat{p}_w = n_{\text{grow}}/n_{\text{tot}}$$

To obtain likelihood-based confidence intervals, we use the test statistic:

$$D(p_w) = 2(\log \mathcal{L}(\hat{p}_w) - \log \mathcal{L}(p_w))$$

i.e. twice the difference in log likelihood between the MLE and any test value of p_w, and solve for the boundaries p_w^* such that $D(p_w^*) = D^*$, the critical value for a chosen significance level in the chi-squared test with one degree of freedom. The MLE and confidence interval boundaries for p_w can simply be transformed to those for α using $\alpha = -\log(1-p_w)$.

Pooling dilution factors: Recall that under the null model of inoculum size effect, and assuming perfect dilution steps, we have

$$\alpha(\bar{N}_i, x) = \alpha(\bar{N}^*, x)/(d_i/d^*)$$

where d_i is the i^{th} dilution factor, normalized by a chosen dilution factor d^*, taken as fixed values. This leaves a single parameter $\alpha^*(x) := \alpha(\bar{N}^*, x)$, the mean number of established cells scaled to the chosen dilution factor, to be estimated in each environment by pooling data across all dilution factors. We thus optimize the likelihood:

$$\log \mathcal{L}_{\text{pool}}\left(\alpha^*(x) | \{(n_{\text{grow}}, n_{\text{tot}})\}_{i=1}^m\right) = \sum_{i=1}^m \log \mathcal{L}\left(1 - \mathrm{e}^{-\alpha^*(x)/(d_i/d^*)} | (n_{\text{grow}}, n_{\text{tot}})_i\right) \tag{S7}$$

where $\log \mathcal{L}$ is the binomial log likelihood defined in Equation S6. Note that we can have different dilution factors d_i (and a different number of them, m) in each environment x. Using the parameterization in terms of α (Model C), each environment can be treated separately.

Working with relative establishment probability: Recall that we can transform the model from a parameterization in terms of $\{\alpha(\bar{N}, x_j)\}$ in all environments x_j, to one in terms of $\alpha(\bar{N}, 0)$ in the baseline environment ($x = 0$) along with $\{\tilde{p}_c(x_j)\}$ in environments $x_j \neq 0$. To obtain MLEs on $\tilde{p}_c$, we could simply substitute MLEs from the original parameterization in terms of α: $\hat{\tilde{p}}_c(x) = \hat{\alpha}(x)/\hat{\alpha}(0)$. However, the confidence intervals must take into account the uncertainty in both the numerator and denominator. This requires running the inference jointly across multiple environments.

We write the joint log likelihood across all conditions, given the data n_{grow} and n_{tot} as arrays indexed by inoculum size (i) and environment (j), as:

$$\begin{aligned}
&\log \mathcal{L}_{\text{joint}}\left(\{\alpha(\bar{N}_i, 0), \{\tilde{p}_c^{(i)}(x_j)\}_{j=1}^s\}_{i=1}^m | n_{\text{grow}}, n_{\text{tot}}\right) \\
&\quad = \sum_{i=1}^m \left[\log \mathcal{L}(1 - \mathrm{e}^{-\alpha(\bar{N}_i, 0)}) + \sum_{j=1}^s \log \mathcal{L}\left(1 - \mathrm{e}^{-\alpha(\bar{N}_i, 0)\tilde{p}_c^{(i)}(x_j)}\right)\right]
\end{aligned} \tag{S8}$$

Proceeding further depends on the particular model (cf. Fig. S1):

- Model A′ (full model): We estimate $\alpha(\bar{N}_i, 0)$ and $\tilde{p}_c^{(i)}(x_j)$ for each inoculum size i separately. For any given value of $\alpha(\bar{N}_i, 0)$, we can simply rescale all $\tilde{p}_c^{(i)}(x_j)$ to obtain the same optimal likelihood. Therefore the estimate of $\alpha(\bar{N}_i, 0)$ can be computed for the single condition $(\bar{N}_i, 0)$ in isolation. However, estimating each $\tilde{p}_c^{(i)}(x_j)$ requires simultaneous consideration of the baseline environment (0) and the focal environment (x_j) at the i^{th} inoculum size.

- Model B′ (fixed environmental effect): We estimate $\alpha(\bar{N}_i, 0)$ for each i, but assume that $\tilde{p}_c(x_j)$ is the same for every i. Adjusting the value of $\tilde{p}_c(x_{j^*})$ in any single environment x_{j^*} affects the optimized values of $\alpha(\bar{N}_i, x_0)$ for all i, which in turn affects the optimized values of all other

$\tilde{p}_c(x_j)$. Thus, we must conduct joint inference on all parameters using the entire data set (i.e. all inoculum sizes and environments) simultaneously.

- Model C′ (null model of inoculum size effect): We estimate a single parameter $\alpha_0^* := \alpha(\bar{N}^*, 0)$ in the baseline environment at some normalizing dilution factor d^*, under the assumption that $\alpha(\bar{N}_i, 0) = \alpha_0^*/(d_i/d^*)$; as well as $\tilde{p}_c(x_j)$ in each non-baseline environment. We can again estimate α_0^* for the baseline environment in isolation (since all $\tilde{p}_c(x_j)$ can be adjusted to obtain the same likelihood in environment x_j for any value of α_0^*). We can estimate $\tilde{p}_c(x_j)$ for each environment x_j using only the data from the baseline and single focal (j^{th}) environments. As in Model C, we do not require the same dilution factors to be used in each environment, but since the data in the baseline and focal environments are used simultaneously, the dilution factors should be normalized by the same factor d^* in both.

When likelihood is a function of more than one parameter, we use the concept of profile likelihood confidence intervals. The profile likelihood function of a given focal parameter is defined as the likelihood when holding this parameter fixed to a given value. The focal parameter's confidence interval is in turn defined by the limits of its fixed values that allow its profile likelihood, optimized over all other parameters, to attain an optimum within a critical difference below the maximum likelihood, as optimized over all parameters including the focal. The critical difference is defined by a chi-squared test with one degree of freedom, since we have fixed one parameter in the profile likelihood.

Model selection: To compare nested models, we use the likelihood ratio test. The test statistic is the deviance (D), equal to twice the difference in maximized log likelihood between the more complex model and the simpler model (i.e. the nested one with fewer parameters). Significance of the deviance is evaluated using a chi-squared test with degrees of freedom equal to the difference in number of model parameters; if non-significant, we accept the simpler model.

Code: Likelihood calculations and numerical optimization were implemented using custom scripts in R, version 3.3.1 (The R Foundation for Statistical Computing, 2016).

10 Generalized linear model for analyzing seeding experiments

10.1 Model description and interpretation

To assess the impact of experimental variables on the seeding experiments, particularly the significance of the streptomycin concentration effect, we fit a generalized linear model (GLM) to observed population growth in wells, treated as binomial data. The fitting is carried out using the built-in R

function 'glm'. The response variable (probability of well growth, p_w) is modelled as a function of the following explanatory variables:

- natural logarithm of the dilution factor applied to the inoculating culture (treated as either continuous or categorical)
- streptomycin concentration on the treatment plates (categorical)
- experiment date when pooling data from more than one experiment (categorical)

We use a complementary log-log (cloglog) link function. The linear predictor then takes the form:

$$\text{cloglog}(p_w) = X\beta$$

where X contains the explanatory variables (i.e. the numerical value of any continuous variable and an indicator for each categorical variable) and β contains the coefficients to be fit.

The rationale for choosing the cloglog link function is that it gives a clean relationship to our theoretical model (Section 8). Under the basic assumption that the number of established cells per well (α) is Poisson-distributed, we have

$$p_w = 1 - \mathrm{e}^{-\alpha}$$

which leads to the simple relationship

$$\text{cloglog}(p_w) := \log(-\log(1 - p_w)) = \log(\alpha)$$

If we further suppose that the number of cells inoculated per well is Poisson-distributed with mean $\bar{N}$ and each cell independently has probability p_c of establishing a surviving lineage (i.e. Model C, the null model of inoculum size effect), we can substitute $\alpha = \bar{N}p_c$ and thus

$$\text{cloglog}(p_w) = \log(\bar{N}) + \log(p_c)$$

such that any explanatory variables expected to affect only inoculum size (i.e. the dilution factor applied to the inoculating culture) are separated from those expected to affect only establishment probability (i.e. streptomycin concentration) in the linear predictor. Specifically, suppose that in the n^{th} observation, $\bar{N} = \bar{N}^*/d_n$ where $\bar{N}^*$ is a baseline culture density and d_n is the normalized dilution factor. Further write $p_{c,n} = p_0 \cdot \tilde{p}_c(x_n)$ where p_0 is the per-cell establishment probability in baseline conditions and $\tilde{p}_c$ is the relative establishment probability (cf. Equation S5) in streptomycin concentration x_n. Then we have

$$\text{cloglog}(p_w(n)) = \log(\bar{N}^*) + \log(p_0) - \log(d_n) + \log(\tilde{p}_c(x_n))$$

This leads to the prediction that, if we treat the logarithm of the dilution factor as a continuous explanatory variable, we expect a fitted coefficient close to -1. The baseline culture density (which may for instance vary from experiment to experiment) and the absolute establishment probability in baseline conditions (which is unknown) will be incorporated into the intercept term of the model fit.

On these theoretical grounds, we treat the logarithm of dilution factor as a continuous explanatory variable in our main analysis. In support of the theoretical model, we indeed find a fitted coefficient close to -1 (see Section 12.2). However, we also conduct an analysis in which dilution factor is treated as a categorical variable, and thus remove any assumptions derived from the theoretical model above. Our conclusions are robust to this choice (see Section 12.2).

Part III

Additional Statistical Results

11 Colony-forming units are Poisson-distributed

We tested whether the number of CFUs in a small volume of highly diluted culture can be described by a Poisson distribution, using colony counts in plated 4μl spots (Section 1). A Poisson distribution was fit using the sample mean. To conduct a goodness-of-fit test, the data (colony count in each spot) were grouped into categories according to the guideline that there should be an expected number of at least five observations per category (ref.[3] p. 540). Deviation from the Poisson distribution was determined at a 5% significance level using a chi-squared test with degrees of freedom equal to the number of categories minus two[3]. We carried out two separate experiments, each with 144 plated spots. In both cases, we could use categories of 0, 1, 2, 3, 4, and 5 or more colonies per spot.

- Experiment 1 (Suppl. Fig. 1a): The sample mean is 1.97 (and variance is 1.68). According to the goodness-of-fit test, the deviation from a Poisson distribution is not significant (χ^2_4: $p = 0.10$) and thus we accept the null model that a Poisson distribution is sufficient to describe the data.
- Experiment 2 (Suppl. Fig. 1b): The sample mean is 1.96 (and variance is 1.77). We again accept the null model (χ^2_4: $p = 0.73$).

12 Seeding experiments

12.1 Theoretical model fitting

Here we report the complete results of fitting our theoretical models of population growth (Sections 8-9) to seeding experiment data. Recall that Model A (equivalently, Model A$'$ in the transformed

parameterization) is the full model, with the number of estimated parameters equal to the number of tested conditions: (# inoculating dilution factors) × (# streptomycin concentrations) = 15 in our experiments. Model B′ allows a separate estimate of mean number of established cells in streptomycin-free media ($\alpha(0)$) at each dilution factor, but assumes relative establishment probability ($\tilde{p}_c(x)$) at each streptomycin concentration ($x \neq 0$) is common to all dilution factors, for a total number of parameters equal to (# dil. factors) + (# non-zero Strep. conc.) = 7 in our experiments. Finally, Model C (or C′), the null model of the inoculum size effect, additionally assumes that the number of established cells scales proportional to inoculum size (inversely proportional to inoculating dilution factor), leaving 1 + (# non-zero Strep. conc.) = 5 parameters to estimate. We fit each model to the data, i.e. number of replicate cultures showing growth, using maximum likelihood estimation of the parameters. Since these models are nested, we then use the likelihood ratio test (LRT) to compare their fits to the data (Section 9). If the deviance (D) between the more complex and the simpler nested model is non-significant, we accept the simpler model. (We sum the contributions to D from each tested streptomycin concentration here, to get an overall model choice for the entire data set.) We use the estimates of $\tilde{p}_c$ from the best-fitting model as selected by the LRT for plotting (main Fig. 3) and comparisons among experiments (Suppl. Table 1). However, the maximum likelihood estimate of $\tilde{p}_c$ tended to be quite consistent among models.

Experiment 1:

- Model A′ vs. B′: $D = 3.56$, $p = 0.89 \Rightarrow$ accept B′
- Model B′ vs. C′: $D = 15.8$, $p = 3.6e-4 \Rightarrow$ reject C′

Further analysis suggests that one of the three inoculating dilution factors (4e7-fold) primarily contributed to the deviation of Model C′ in this experiment, suggesting an inaccuracy in preparation of this particular dilution that resulted in lack of proportionality in inoculum sizes as assumed by this model. Indeed, if the data from this dilution factor are excluded from the analysis, Model C′ is accepted ($D = 0.210$, $p = 0.65$ compared to Model B′). For comparison, dropping either of the other single dilution factors still results in rejection of Model C′, suggesting that the test is not simply under-powered on the reduced data set, but rather that the identified dilution factor is the source of error. Nonetheless, we obtain similar point estimates for $\tilde{p}_c$ either with or without this dilution factor; for consistency, we report results including all dilution factors.

Experiment 2:

- Model A′ vs. B′: $D = 8.99$, $p = 0.34 \Rightarrow$ accept B′
- Model B′ vs. C′: $D = 1.65$, $p = 0.44 \Rightarrow$ accept C′

Thus, in this experiment, we accept the Model C′ as the best fit.

12.2 Generalized linear model fitting

Here we report the complete results of fitting the generalized linear model, as described in Section 10, to population growth data from the seeding experiments. Recall that the explanatory variables are streptomycin concentration, [Strep] for short; the logarithm of the inoculating dilution factor, log(dilfac) for short; and experiment date, if applicable.

Experiment 1: Fitting a full model identified that the log(dilfac) and [Strep] main effects were significant, but their interaction was not. A reduced model including only the main effects was correspondingly preferred by the Akaike Information Criterion (AIC: 90.2 for the full model vs. 83.7 for the reduced model if log(dilfac) is treated as a continuous variable, or 96.1 vs. 83.7 if treated as a categorical variable). This result is in qualitative agreement with our theoretical model fit accepting Model B′, in which the effect of inoculation density is separated from that of relative establishment probability (the latter expected to depend only on streptomycin concentration). The reduced model fit indicated that the effect of [Strep] at $1/16\times\mathrm{MIC}_R$ and $1/8\times\mathrm{MIC}_R$ is significant relative to the streptomycin-free conditions ($p = 0.01$ and $p <$ 2e-16, respectively, regardless of whether log(dilfac) is treated as continuous or categorical). As expected, the effect of log(dilfac) is also highly significant ($p <$2e-16 when treated as continuous); here, the fitted coefficient of -1.30 is reasonably close to the theoretical prediction of -1 (see Section 10), but likely skewed by the suspected error in one dilution factor as mentioned above. The effect of log(dilfac) remains highly significant when treated as categorical (lowest vs. highest dilution factor: $p <$2e-16; middle vs. highest: p=3e-10).

Experiment 2: Again, fitting the full model with all explanatory variables indicated that the log(dilfac) and [Strep] main effects were significant, but their interaction was not; the reduced model excluding the interaction term was preferred by AIC (86.5 vs. 81.6 with log(dilfac) as continuous; 89.0 vs. 82.0 with log(dilfac) as categorical). The reduced model fit again indicated that [Strep] was significant at $1/16\times\mathrm{MIC}_R$ and at $1/8\times\mathrm{MIC}_R$ (p=2e-7 and $p <$2e-16, respectively, regardless of whether log(dilfac) is treated as continuous or categorical). The effect of log(dilfac) is also highly significant ($p <$2e-16 in all cases), and when treated as continuous, the fitted coefficient of -1.01 is in excellent agreement with the theoretical prediction, consistent with the acceptance of the theoretical Model C′ in the previous analysis.

Pooling both experiments: Using experiment date as an additional explanatory variable, a hierarchical search (using the built-in function 'step' on the fitted full model) according to minimal AIC identified a reduced model in which all main effects and the experiment date × log(dilfac) interaction

effect are retained. This interaction term presumably arises because of the suspected inaccuracy in the lowest dilution factor (4e7-fold) only in Experiment 1, as described above. Indeed, the interaction is identified as significant ($p = 0.003$) with log(dilfac) taken as continuous, while with log(dilfac) taken as categorical with the highest dilution factor (1.6e8-fold) as the baseline, the interaction between experiment date and the lowest dilution factor (4e7-fold) is significant ($p = 0.008$) but the interaction with the middle dilution factor (8e7-fold) is not significant ($p = 0.8$). Taken together, these results point to error in a single dilution step as the source of deviating effects; we do not interpret this result as having biological significance. More importantly, pooling data from two experiments strengthens the conclusions regarding the effect of streptomycin: specifically, [Strep] at $1/16\times\mathrm{MIC}_R$ or $1/8\times\mathrm{MIC}_R$ has a highly significant effect relative to the streptomycin-free control (p=2e-8 and p <2e-16, respectively, regardless of whether log(dilfac) is treated as continuous or categorical).

13 Testing the null model of the inoculum size effect

Here we report the complete results of fitting our theoretical models of population growth (see Sections 8-9) to population growth data, where we are interested in testing the null model for the effect of inoculum size at a given streptomycin concentration. Specifically, at each streptomycin concentration, we test whether the null model (Model C or equivalently C$'$) is accepted, by the likelihood ratio test, in comparison to the full model (Model A or A$'$) that allows an arbitrary effect of each separate inoculum size. The results of the "main experiment" (testing both $1/16\times\mathrm{MIC}_R$ and $1/8\times\mathrm{MIC}_R$ streptomycin, as well as streptomycin-free conditions) are illustrated in main Fig. 5 and Suppl. Fig. 5; the results of the two "supplementary experiments" (one testing $1/16\times\mathrm{MIC}_R$ and one testing $1/8\times\mathrm{MIC}_R$ streptomycin, along with a streptomycin-free control in each) are illustrated in Suppl. Fig. 6; and estimates of relative establishment probability $\tilde{p}_c$ from all experiments are summarized in Suppl. Table 1. In Table S2 we summarize the results of the likelihood ratio test at each streptomycin concentration in each experiment. In every case, the deviance (D) of the null model from the full model is non-significant ($p > 0.05$), and thus we accept the null model.

Table S2: Testing the null model of inoculum size effect: likelihood ratio test results

Experiment	[Strep] $\times \mathrm{MIC}_R$	# inoc. sizes tested	Deviance* (null from full)	p-value**	Effective O/N culture density† (viable cells/ml)
main	0	5	3.08	0.55	7.58×10^9
	1/16	9	9.74	0.28	
	1/8	6	2.91	0.71	
suppl. 1	0	5	8.57	0.073	7.17×10^9
	1/16	9	1.02	1.00	
suppl. 2	0	5	0.347	0.99	9.10×10^9
	1/8	10	13.1	0.16	

* Deviance (D) of the null model from the full model, defined as twice the difference in maximal log likelihood of each model (see Section 9).

** p-value in the likelihood ratio test, i.e. a chi-squared test with D as the test statistic and degrees of freedom = (# parameters in full model – # parameters in nested model) = (# tested inoculum sizes – 1).

† In streptomycin-free conditions, we use the maximum likelihood estimate of $\alpha(0) \equiv \bar{N}_{\mathrm{eff}}$, scaled up by the corresponding dilution factor applied for the inoculation, to estimate an effective viable cell density in the overnight culture used for inoculation. This is equivalent to the "most probable number" method for determining bacterial density using multiple dilution factors [2]. We use this estimate to calibrate the "effective mean inoculum size" in the x-axis of our plots.

References

[1] European Committee for Antimicrobial Susceptibility Testing (EUCAST). Determination of minimum inhibitory concentrations (MICs) of antibacterial agents by broth dilution. Clin. Microbiol. Infect., 9:1–7, 2003.

[2] M. A. Hurley and M. E. Roscoe. Automated statistical analysis of microbial enumeration by dilution series. J. Appl. Bacteriol., 55:159–164, 1983.

[3] R. J. Larsen and M. L. Marx. An Introduction to Mathematical Statistics and Its Applications. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 3rd edition, 2001.

[4] I. Levin-Reisman, I. Ronin, O. Gefen, I. Braniss, N. Shoresh, and N. Q. Balaban. Antibiotic tolerance facilitates the evolution of resistance. Science, 355:826–830, 2017.

[5] G. Martin, R. Aguilée, J. Ramsayer, O. Kaltz, and O. Ronce. The probability of evolutionary rescue: towards a quantitative comparison between theory and evolution experiments. Phil. Trans. R. Soc. B, 368:20120088, 2013.

[6] D. Schleheck, N. Barraud, J. Klebensberger, J. S. Webb, D. McDougald, S. A. Rice, and S. Kjelleberg. *Pseudomonas aeruginosa* PAO1 preferentially grows as aggregates in liquid batch cultures and disperses upon starvation. PLoS One, 4:e5513, 2009.