

Supplementary Text 1: Breakpoint and the minimum inhibitory concentration (MIC)

Conceptually, a breakpoint is the threshold concentration of an antibiotic used to define whether a bacterial strain is likely to be treatable by a given antibiotic or not in a patient. However, the process for setting breakpoints is often unclear. As described in the reference [29], breakpoint setting requires “the integration of knowledge of the wild-type distribution of MICs, assessment of the pharmacokinetics/pharmacodynamics of the antibacterial, and study of the clinical outcome of infections when the antibacterial is used.” Notably, two major breakpoint-setting organizations, the European Committee on Antimicrobial Susceptibility Testing (EUCAST) and the Clinical and Laboratory Standards Institute (CLSI) recommend different breakpoint values for some antibiotics, indicating there is not a single straightforward procedure to determine the breakpoint values. The values can change over time depending on the clinical trials/outcomes. The minimum inhibitory concentration (MIC) represents the concentration at which bacterial growth and death are balanced, such that the population size as a whole ceases to change. In practice, the relationship between the MIC and the breakpoint is that a bacterial strain is categorized as susceptible if $MIC < \text{breakpoint}$ and resistant if $MIC > \text{breakpoint}$.

Supplementary Text 2: Mathematical description of population dynamics.

We describe our model development in three different sections, increasing in complexity. The first section considers cell reproduction. The second section includes cell death. The third section describes reproduction and death of cells with two distinct phenotypes.

1) Cell reproduction. In our study, we defined live cells as those cells that can give birth. Denoting the number of live cells at time t by N_t and the number of offspring from a live cell by B , the number of live cells at the next generation time $t+1$ is

$$N_{t+1} = BN_t. \quad \text{Eq. 1}$$

2) Cell reproduction and death. In the presence of antibiotics, cells can be killed by antibiotics. Defining the probability that a cell dies as K ,

$$N_{t+1} = (1 - K)BN_t. \quad \text{Eq. 2}$$

Then, the number of cells killed is

$$F_{t+1} = KBN_t. \quad \text{Eq. 3}$$

Eqs 2 and 3 can be written in matrix form,

$$\begin{bmatrix} N_{t+1} \\ F_{t+1} \end{bmatrix} = B \begin{bmatrix} (1 - K) & 0 \\ K & 0 \end{bmatrix} \cdot \begin{bmatrix} N_t \\ F_t \end{bmatrix}. \quad \text{Eq. 4}$$

This matrix is a population projection matrix, whose eigenvector and eigenvalue of these matrixes represent the equilibrium distribution of a population and steady-state growth rate of the population¹.

To calculate the eigenvalue for the matrix,

$$B \begin{bmatrix} (1-K) & 0 \\ K & 0 \end{bmatrix} \cdot \begin{bmatrix} N_t \\ F_t \end{bmatrix} = \lambda \begin{bmatrix} N_t \\ F_t \end{bmatrix}. \quad \text{Eq. 5}$$

Using the characteristic equation above, we can determine the eigenvalue of the matrix, λ

$$\lambda = (1-K)B, \quad \text{Eq. 6}$$

which corresponds to the growth rate of a population.

Of course, this growth rate is proportional to the number of offspring B . If you divide it by B , we have

$$\mu = \frac{\lambda}{B}, \quad \text{Eq. 7}$$

which indicates the probability that a cell would divide. In our analysis with an antibiotic, this corresponds to the survival probability (because we defined cell survival as whether a cell endured the antibiotic exposure and divided). This interpretation agrees with its mathematical expression. Because K is the probability that a cell is killed (i.e., failed to divide), $\mu = (1-K)$ should be equal to the probability that a cell survives. Please note that a cell survived antibiotic exposure and divide will give birth to two daughter cells. Only when these two daughter cells divide (Supplementary Fig. 2a), the survival of the original cell is considered robust survival.

3) Reproduction and death of cells in our model. Now, we categorize live cells into two states: healthy (H) and vulnerable (V) states. The details of these states were described in the main text. Briefly, H or V cells can divide and give birth. P defines the probability that a daughter cell from a H mother cell is in the V state. Q defines the probability that a daughter cell from a V mother cell is in the H state. Here, two daughter cells take on their states independently with each other.

$$H_{t+1} = B(H_t - PH_t + QV_t) \quad \text{Eq. 8}$$

$$V_{t+1} = B(PH_t + V_t - QV_t - K)V_t \quad \text{Eq. 9}$$

$$F_{t+1} = BKV_t \quad \text{Eq. 10}$$

Rewriting Eqs 8-9 in matrix form,

$$\begin{bmatrix} H_{t+1} \\ V_{t+1} \\ F_{t+1} \end{bmatrix} = B \begin{bmatrix} 1-P & Q & 0 \\ P & 1-Q-K & 0 \\ 0 & K & 0 \end{bmatrix} \begin{bmatrix} H_t \\ V_t \\ F_t \end{bmatrix}. \quad \text{Eq. 11}$$

Its eigenvector is

$$\begin{bmatrix} H_s \\ V_s \\ F_s \end{bmatrix} = B \begin{bmatrix} \frac{1}{4KP} (\sqrt{2Q(K+P) + (K-P)^2 + Q^2} + K - P + Q) (\sqrt{2Q(K+P) + (K-P)^2 + Q^2} - K - P - Q + 2) \\ \frac{1}{2K} (\sqrt{2Q(K+P) + (K-P)^2 + Q^2} - K - P - Q + 2) \\ 1 \end{bmatrix}.$$

Eq. 12

The eigenvalue, λ , can be calculated. Dividing it by B , as discussed Eq. 7, we have.

$$\mu = \frac{1}{2}(\sqrt{2Q(K+P) + (K-P)^2 + Q^2} - K - P - Q + 2) \quad \text{Eq. 13}$$

This provides a constraint for K , P , Q , and μ . For example, we can rewrite this equation to describe K as a function of P , Q , and μ values by:

$$K = \frac{PQ}{\mu+P-1} + 1 - Q - \mu, \quad \text{Eq. 14}$$

The μ in Eq. 13 denotes a probability of cell division (as discussed in Eq. 7 and below). Here, the mathematical expression of K is no longer equal to $1 - \mu$ (see section 2) because we assume that only cells in one state (V) die. In our experiments, we aimed to achieve steady growth during the microscope experiment by pre-treating cells with cefsulodin in liquid culture. Indeed, in our previously published work ², where the same protocol was used to track the total number of live and dead cells under cefsulodin exposure, we found that the number of live cells increases exponentially over time (Figure Supplement 5A in ref ²). The numbers of dead cells and live cells show a linear relationship (Figure Supplement 5B). These observations support the steady state growth.

Finding Kinship Survival Correlations

Probability distribution of cells in different cellular states

We will now expand this model to find kinship survival correlations. When a cell divides, the lineage splits into two. When a cell dies, the lineage ends. However, these two different lineage configurations make it challenging to compute the correlation. To address this issue, we consider another state: *never-born* (NB). With this new dummy-state, every cell produces “offspring,” including dead cells, which produces B never-born cells. Never-born cells also produce B never-born cells. However, this never-born state is merely a place-holder to maintain the lineage structure. We distinguish between spots on the family tree where the cell died and spots where there was never a cell to begin with. The former will be considered in the calculation of correlation, but not the latter.

Denoting the number of never-born cells by U , we have

$$U_{t+1} = BF_t + BU_t. \quad \text{Eq. 15}$$

With this consideration, we now have a 4×4 matrix:

$$\begin{bmatrix} H_{t+1} \\ V_{t+1} \\ F_{t+1} \\ U_{t+1} \end{bmatrix} = B \begin{bmatrix} 1-P & Q & 0 & 0 \\ P & 1-Q-K & 0 & 0 \\ 0 & K & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} H_t \\ V_t \\ F_t \\ U_t \end{bmatrix}. \quad \text{Eq. 16}$$

Alternatively, defining the projection matrix as T ,

$$\mathbf{T} = \begin{bmatrix} 1-P & Q & 0 & 0 \\ P & 1-Q-K & 0 & 0 \\ 0 & K & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad \text{Eq. 17}$$

Eq. 16 can be written as

$$\begin{bmatrix} H_{t+1} \\ V_{t+1} \\ F_{t+1} \\ U_{t+1} \end{bmatrix} = B\mathbf{T} \begin{bmatrix} H_t \\ V_t \\ F_t \\ U_t \end{bmatrix}. \quad \text{Eq. 18}$$

Now, we apply this equation to a progenitor cell. For example, consider that this progenitor is in a healthy state. When it gives birth, the number distribution of healthy, vulnerable, dead and never-born cells in the offspring can be calculated by

$$\begin{bmatrix} H_1 \\ V_1 \\ F_1 \\ U_1 \end{bmatrix} = B\mathbf{T} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad \text{Eq. 19}$$

The total number of offspring, $H_1 + V_1 + F_1 + U_1$, is equal to B .

Alternatively, we can simply divide the number distribution of cells by B to get the *probability* distribution of healthy, vulnerable, dead and never-born cells (instead of cell numbers), i.e.,

$$\begin{bmatrix} p_{H,1} \\ p_{V,1} \\ p_{F,1} \\ p_{U,1} \end{bmatrix} = \mathbf{T} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad \text{Eq. 20}$$

The column vector on the left-side describes the *1st* descendants from a healthy progenitor cell, in particular, their *probability* distribution in different cellular states.

We can further expand this operation to find the probability distribution for the i^{th} descendent, raising $\mathbf{T}$ to the i^{th} power,

$$\begin{bmatrix} p_{H,i} \\ p_{V,i} \\ p_{F,i} \\ p_{U,i} \end{bmatrix} = \mathbf{T}^i \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad \text{Eq. 21}$$

Likewise, we can calculate the probability distribution of the cellular states in the i^{th} descendent from a *vulnerable* progenitor cell,

$$\begin{bmatrix} p_{H,i} \\ p_{V,i} \\ p_{F,i} \\ p_{U,i} \end{bmatrix} = \mathbf{T}^i \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}. \quad \text{Eq. 22}$$

Now, we will apply this matrix to two cells from a common progenitor, whose distance in the relationship is represented by $[i, j]$; see Fig. 2a. These cells can either 1) survive antibiotic exposure and is alive or 2) killed by the antibiotic and is dead. The probability that both cells are killed:

$$n' = p_{F,i}p_{F,j}. \quad \text{Eq. 23}$$

Now, we consider the probability that the first cell survives but the second cell is killed. Because cells in healthy and vulnerable states survive, this probability is given by

$$x' = (p_{H,i} + p_{V,i})p_{F,j}. \quad \text{Eq. 24}$$

Likewise, the probability that the first cell was killed and only the second cell survived is

$$y' = p_{F,i}(p_{H,j} + p_{V,j}). \quad \text{Eq. 25}$$

The probability that both cells survive is

$$b' = (p_{H,i} + p_{V,i})(p_{H,j} + p_{V,j}). \quad \text{Eq. 26}$$

As discussed above, the never-born cells do not figure into these calculations.

We can calculate n' , x' , y' , and b' for both healthy and vulnerable progenitors (hence the subscripts for b_H , b_V , etc.) using Eqs 21 and 22 respectively. However, to get the value representative for a population, we need to weight them by the number of healthy and vulnerable progenitors at the steady state by multiplying them by H_s and V_s , which we have described mathematically in Eq. 12. With this consideration, we now have

$$n = \frac{1}{A_0}(H_s n'_H + V_s n'_V), \quad \text{Eq. 27}$$

$$x = \frac{1}{A_0}(H_s x'_H + V_s x'_V), \quad \text{Eq. 28}$$

$$y = \frac{1}{A_0}(H_s y'_H + V_s y'_V), \quad \text{Eq. 29}$$

$$b = \frac{1}{A_0}(H_s b'_H + V_s b'_V), \quad \text{Eq. 30}$$

Where A_0 is a normalization factor such that n , x , y , and b add to 1.

$$A_0 = H_s(b'_H + x'_H + y'_H + n'_H) + V_s(b'_V + x'_V + y'_V + n'_V). \quad \text{Eq. 31}$$

Pearson Correlation Coefficient

With these quantities, we can calculate the Pearson correlation coefficient. The derivation is as follows, starting from the coefficient's definition (here, z and w are simply two related datasets, in this case survival values for the first and second cells in each pair; $\bar{z}$ and $\bar{w}$ are their respective means, i is a counter and N is the number of datapoints, with each datapoint being a pair of cells with the selected relationship):

$$C_{zw} = \frac{\sum_{i=1}^N (z_i - \bar{z})(w_i - \bar{w})}{\sqrt{\sum_{i=1}^N (z_i - \bar{z})^2} \sqrt{\sum_{i=1}^N (w_i - \bar{w})^2}} = \frac{\sum_i z_i w_i - N \bar{z} \bar{w}}{\sqrt{\sum_i z_i^2 - N \bar{z}^2} \sqrt{\sum_i w_i^2 - N \bar{w}^2}} \quad \text{Eq. 32}$$

For a binary dataset, in which all individual datapoints are ones and zeros, each term in Eq. 32 can be written in terms of the four probabilities n , b , x and y :

$$\sum_i z_i w_i = Nb, \quad \text{Eq. 33}$$

$$\sum_i z_i^2 = N \bar{z} = N(b + x). \quad \text{Eq. 34}$$

Here, $\sum_i z_i^2 = N \bar{z}$, because z is either 1 or 0.

$$\bar{z}^2 = b^2 + 2xb + x^2. \quad \text{Eq. 35}$$

$$\sum_i w_i^2 = N \bar{w} = N(b + y) \quad \text{Eq. 36}$$

$$\bar{w}^2 = b^2 + 2yb + y^2 \quad \text{Eq. 37}$$

These equivalencies can be inserted into the correlation equation to get

$$C = \frac{Nb - N(b+x)(b+y)}{\sqrt{N(b+x-b^2-2xb-x^2)} \sqrt{N(b+y-b^2-2yb-y^2)}}, \quad \text{Eq. 38}$$

which simplifies to

$$C = \frac{b-bx-by-xy-b^2}{\sqrt{(x+b)(1-x-b)(y+b)(1-y-b)}}. \quad \text{Eq. 39}$$

Summarizing the key steps in our mathematical derivation, we defined P , Q , and K , as the damage rate, the recovery rate, and death rates. We then determined the probability distribution of cells in different cellular states using Eq. 21-22, and applied it to pairs of cells to calculate the probability for different combinations of cell fates (Eq. 27-30). We then apply it to determine the correlation coefficient C using Eq. 39 for any $[i, j]$ relationship.

In our work, we varied the P , Q , and K values, calculated C for each parameter set using Eq. 39, and fit it to the experimentally determined C (Fig. 2bc), thereby identifying the P , Q , and K values that best explain the data. The algorithm searched for the best-fitting parameters in a square where P and Q ranged from 0 to 1 (increasing by 0.02, thereby 2500 different combinations of P and Q values). K was fixed by P , Q , and the measured μ via Eq. 14. We compared the calculated C (which we denote by C_{cal} going forward) to the experimentally measured C for each parameter set. The χ^2 value $[= \sum \frac{(C - C_{\text{cal}})^2}{C}]$, where the sum was performed for all nine relationships shown in Fig. 2bc] was calculated to choose the P , Q , K combination that minimizes the χ^2 . We then zoomed in the chosen combination and varied P and Q near this combination and repeated the iteration until we find the best fit to seven decimal places. Please note that when we calculated χ^2 , all nine measured kin correlations were given equal weight, although the number of samples decreases for more distant relationships. Yet, we observed a

good agreement between the experimental C (horizontal bars) and calculated C_{cal} (line) in Fig. 2bc.

Supplementary Text 3: Statistical Tests

3.1) Fig. 2bc: To statistically assess whether the obtained correlation coefficients are significantly greater than zero, i.e., no correlation, we performed linear regression analyses on the survival parameters (X) between related cells and calculated corresponding p-values. As described, three biological replicate experiments were conducted. We computed p-values for each replicate individually as well as for the combined dataset.

Supplementary Table 1: p-values for Fig. 2b wild-type, calculated based on two-sided t-test

Relation		P-value			
		Combined	Rep 1	Rep 2	Rep 3
[1,1]	Sibling	<0.00005	<0.00005	<0.00005	0.0126
[2,1]	Aunt	0.0001	0.0026	<0.00005	0.6122
[1,2]	Niece	0.0002	0.0025	0.0001	0.8689
[2,2]	1 st Cousin	<0.00005	<0.00005	0.034	0.12
[3,1]	Great Aunt	0.0122	0.2628	0.0286	0.2448
[1,3]	Great Niece	0.0282	0.3907	0.0077	0.7432
[2,3]	1 st Cousin 1 removed	0.0151	0.5809	0.1579	0.0358
[3,2]	1 st Cousin 1 removed	0.0041	0.5472	0.0141	0.1321
[3,3]	2 nd Cousin	0.67	0.0367	0.954	0.329

Supplementary Table 2: p-values for Fig. 2c $\Delta tolC$, calculated based on two-sided t-test

Relation		P-value			
		Combined	Rep 1	Rep 2	Rep 3
[1,1]	Sibling	<0.00005	0.0012	0.085	<0.00005
[2,1]	Aunt	<0.00005	0.0001	0.0341	<0.00005
[1,2]	Niece	<0.00005	0.0001	0.0071	<0.00005
[2,2]	1 st Cousin	<0.00005	0.0596	0.0071	<0.00005
[3,1]	Great Aunt	<0.00005	0.5924	0.2501	<0.00005
[1,3]	Great Niece	<0.00005	0.3101	0.2688	<0.00005
[2,3]	1 st Cousin 1 removed	<0.00005	0.0795	0.1727	<0.00005
[3,2]	1 st Cousin 1 removed	<0.00005	0.0386	0.1847	<0.00005
[3,3]	2 nd Cousin	<0.00005	0.3479	0.2303	<0.00005

3.2) Fig. 2d. To test whether the increase of $\langle X \rangle_Z - \langle X \rangle$ with Z is positive and strong, we performed bivariate correlation between $\langle X \rangle_Z - \langle X \rangle$ and Z using the data from all the replicates ($n=21$, open circles in the figure, whose values can be found in Source data 1) and calculated Pearson's r . The r was 0.76 (p-value = 0.00005, based on a t-distribution, two-sided) for WT and 0.64 (p-value= 0.0018) for $\Delta tolC$, showing a strong positive increase and statistical significance in both cases. We next evaluated whether the slope of this increase between WT and $\Delta tolC$ strains is statistically different. We determined the slope for each biological replicate ($n=3$). A

two-tailed heteroscedastic t-test on three slopes shows the p-value = 0.98, meaning the slope of increase between WT and $\Delta tolC$ are not statistically different.

3.3) Fig. 3a: To statistically test whether the observed $F(R_N)/F_0(R_N)$ is significantly greater than 1, we performed a one-sided t-test between $F(R_N)/F_0(R_N)$ and 1 for each N using the weighted means and standard errors (as plotted in Fig. 3a, whose values can be found in Source data 1). See the table below.

Supplementary Table 3: p-values for Fig. 3a based on one-sided t-test

	WT	$\Delta tolC$
N = 1	0.01026	0.208006
N = 2	0.003877	0.013346
N = 3	0.044732	0.031276

To statistically evaluate the increase of $F(R_N)/F_0(R_N)$ with N , we performed bivariate correlation between $F(R_N)/F_0(R_N)$ and N using the data from three replicates (whose values can be found in Source data 2). In some replicates, especially for $N=3$, the sample size was too low (below 10). These replicates were not considered in the correlation analysis: hence $n=7$. Pearson's r was 0.89 (p-value=0.003 based on a t-distribution, one-sided) for WT and 0.82 (p-value=0.011) for $\Delta tolC$, showing a strong positive increase and statistical significance in both strains.

3.4) Fig. 3b: To statistically test whether $P(D = 2 | R_N)$ is significantly greater than $P(D = 2)$, we performed a one-sided t-test between them for each N using the weighted means and standard errors (as plotted in Fig. 3b, whose values can be found in Source data 1). See the table below for the p-value.

Supplementary Table 4: p-values for Fig. 3b based on one-sided t-test

	WT	$\Delta tolC$
N = 1	0.024435	0.078813
N = 2	0.007732	0.017265
N = 3	0.074612	0.002586

To statistically evaluate the increase of $P(D = 2 | R_N)$ with N , we performed bivariate correlation between $P(D = 2 | R_N)$ and N using the data from three replicates (whose values can be found in Source data 2). In some replicates, especially for $N=3$, the sample size was too low (below 10). These replicates were not considered in the correlation analysis: hence ($n=7$). Pearson's r was 0.84 (p-value = 0.009 based on a t-distribution, two-sided) for WT and 0.74 (p-value = 0.028) for $\Delta tolC$, showing a strong positive increase and statistical significance in both strains.

3.5) Fig. 3c: To statistically test whether $P(D = 0 | R_N)$ is significantly less than $P(D = 0)$, we performed a one-sided t-test between them for each N using the weighted means and standard

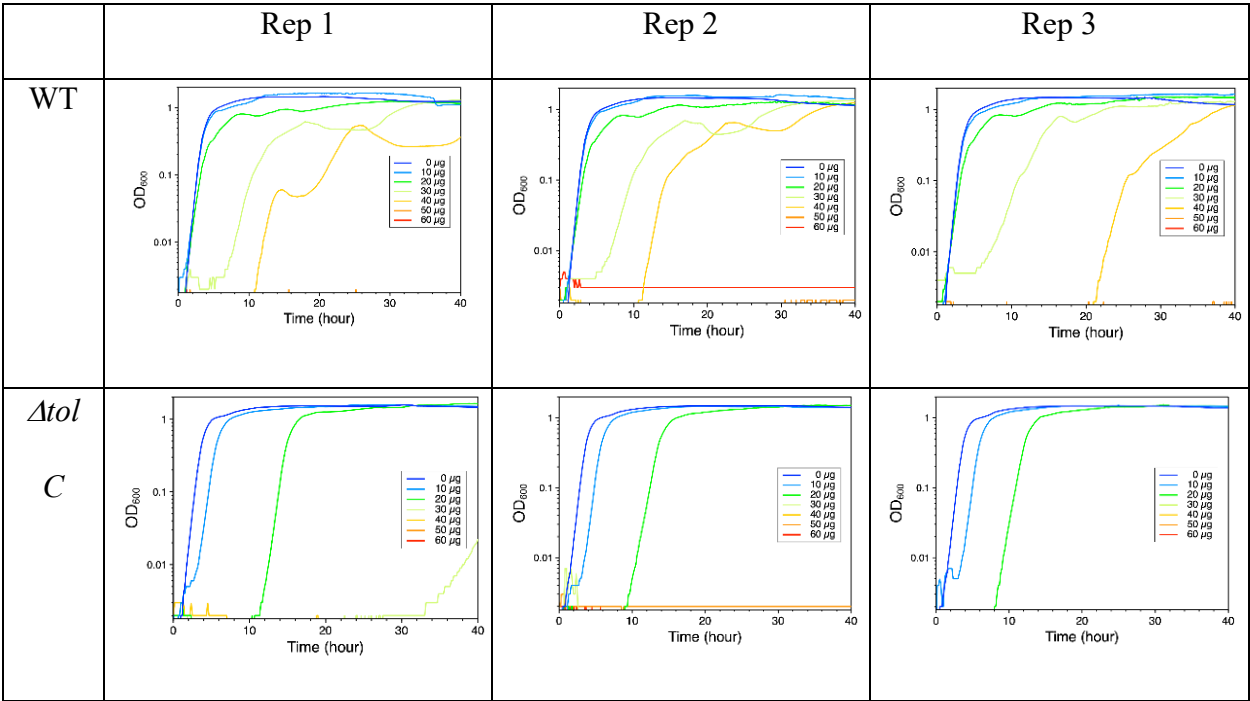
errors (as plotted in Fig. 3c, whose values can be found in Source data.xlsx). See the table below for the p-value.

Supplementary Table 5: p-values for Fig. 3c based on one-sided t-test

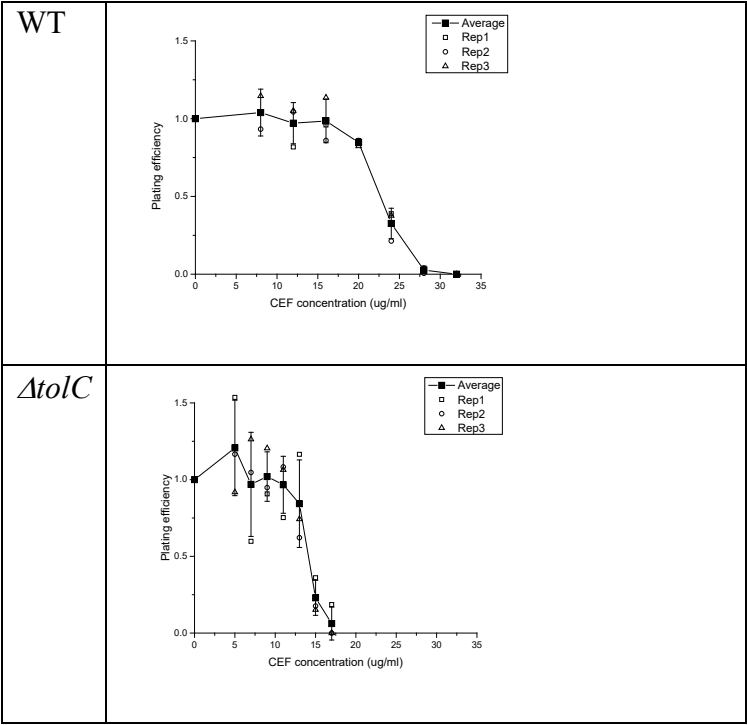
	WT	$\Delta tolC$
N = 1	0.029465	0.245533
N = 2	0.248832	0.065129
N = 3	0.026832	0.006273

To statistically evaluate the decrease of $P(D = 0 | R_N)$ with N , we performed bivariate correlation between $P(D = 0 | R_N)$ and N using the data from the replicates (whose values can be found in Source data 2). In some replicates, especially for $N=3$, the sample size was too low (below 10). These replicates were not considered in the correlation analysis: hence ($n=7$). Pearson's r was -0.91 (p-value=0.002 based on a t-distribution, two-sided) for WT and -0.76 (p-value=0.024) for $\Delta tolC$, showing a strong negative trend and statistical significance in both strains.

a)



b)



Supplementary Fig. 1. Growth curve under different concentrations of cefsulodin.

a) Fig. 1a shows images of WT and $\Delta tolC$ cultures after 24 hours of incubation with different concentrations of cefsulodin. We plotted corresponding growth curves (OD₆₀₀ versus time) in three biological replicates.

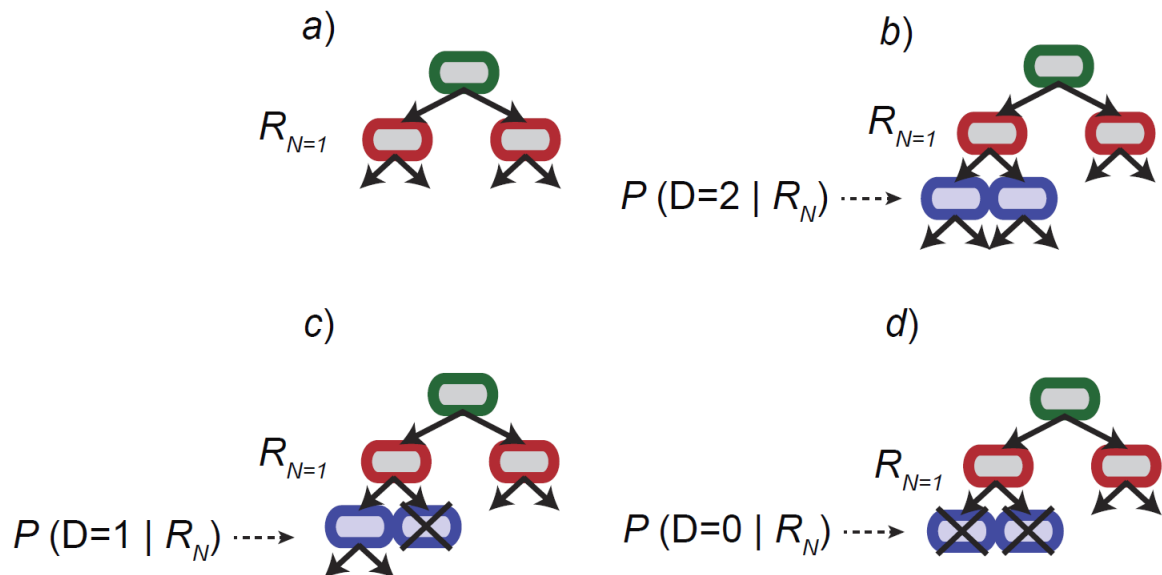
Procedure: Bacteria were grown until mid-log-phase (OD₆₀₀ ~0.4) in a liquid culture. The culture was diluted to the OD₆₀₀ of 0.001. 100 μ L of the culture was transferred to a sterile flat-bottomed 96 well plate. Cells were grown at 37°C while shaking at 567 cpm (3 mm) and OD₆₀₀ was measured every 10 minutes using a BioTek Synergy H1 plate reader. The experiment was repeated three times.

b) Plating efficiency: The CFU assay using agar plates was performed to measure the MIC.

Small open symbols represent values from three replicates. Solid symbols show their average.

The error bars represent their standard deviation.

In a liquid culture, the MIC was 50 μ g/ml for WT (Supplementary Fig. 1a). On agar, the plating efficiency dropped to zero at 32 μ g/ml (Supplementary Fig. 1b), so MIC = 32 μ g/ml, which was slightly lower. We observed a similar pattern for $\Delta tolC$: 30 μ g/ml in liquid and 20 μ g/ml on agar (Supplementary Fig. 1a-b). Please note that we measured the MICs with fine increments of antibiotic concentrations to identify the concentration suitable for our single-cell-level studies. In typical measurements in the field, the concentration is increased by twofold. Therefore, it is common to see a twofold difference in MIC determination. In this regard, the MIC difference between liquid and agar we observed is considered minimal.



e) Table

WT

Number of generations of cells considered	Total number of cells analyzed	Number of robust cells observed, $F(R_N)$	Expected number based on random chance, $F_0(R_N)$
N=1	5595	2548	2437.2
N=2	5271	1211	435.7
N=3	4875	92	14.5

$\Delta tolC$

Number of generations of cells considered	Total number of cells analyzed	Number of robust cells observed, $F(R_N)$	Expected number based on random chance, $F_0(R_N)$
N=1	6796	3015	2960.3
N=2	6517	1666	538.7
N=3	6159	204	18.3

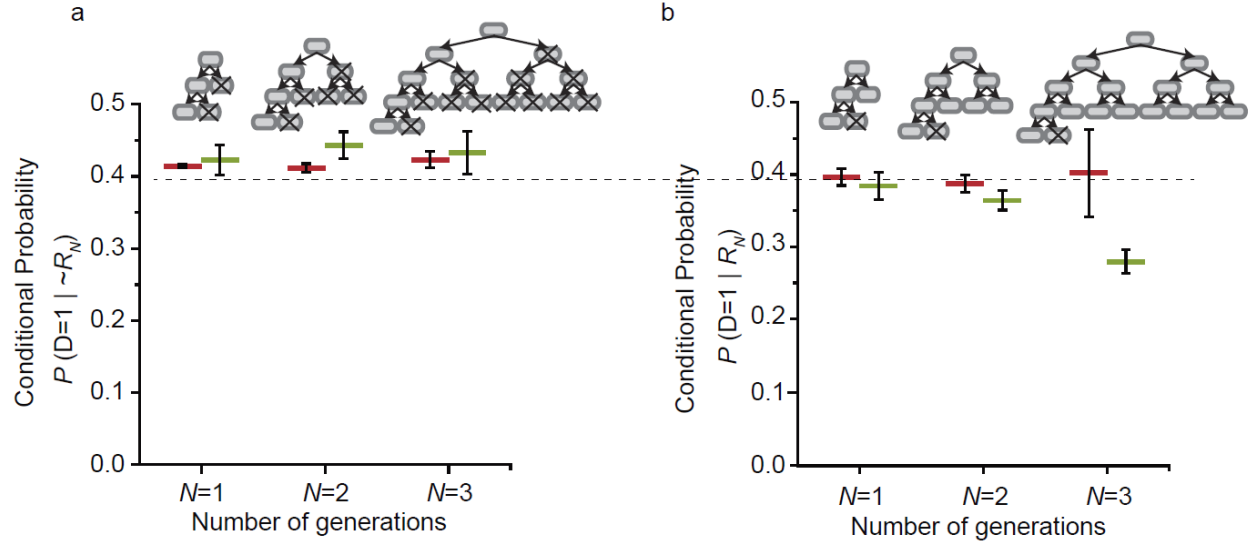
Supplementary Fig. 2. Graphical explanation of robustness and its number (or frequency).

a. A mother cell (marked as green) is alive if it divides. However, this division itself does not make the mother cell robust. For it to be robust, its two daughter cells (red cells) must divide. To clarify this, we explicitly added arrows for daughter cells.

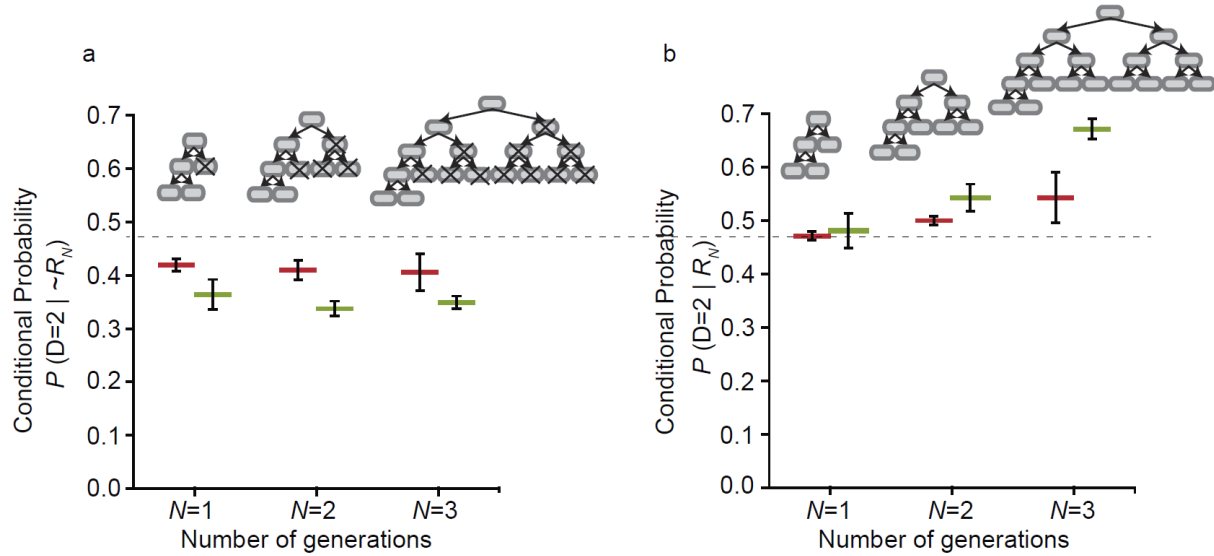
b,c,d. For a daughter cell (a left red cell in panel ‘b’) to be robust, both of its own daughter cells (blue cells) should divide. The D is used to denote the number of blue cells that survive (as denoted by the arrows below them). The panel ‘b’, ‘c’ and ‘d’ respectively denotes the conditional probability $P(D = 2 \mid R_{N=1})$, $P(D = 1 \mid R_{N=1})$, and $P(D = 0 \mid R_{N=1})$.

e). Table: We identified the robust cells in our dataset and determined their frequency: $F(R_N)$. Since we consider the robustness of one generation (only mother cells), N is equal to 1. We then repeated this procedure for two consecutive generations exhibiting robustness, i.e., robust grandmother, mother, and aunt ($N=2$) and for three consecutive generations of robustness ($N=3$).

The expected number of robust lineages based on random chance, $F_0(R_N)$, is calculated as follows. As defined in the main text, the probability that a cell survives is $\langle X \rangle$, which we calculate by analyzing the fate of all cells. Then, the probability that two daughter cells survive is expected to be $\langle X \rangle^2$. Therefore, the expected number of robust cells based on random chance, $F_0(R_{N=1})$, is the total number of cells multiplied by $\langle X \rangle^2$ for $N = 1$. For two consecutive robust generations, we need to consider the survival of the next generation (four additional cells): thus $F_0(R_{N=2}) =$ the total number of cells multiplied by $\langle X \rangle^{2+4}$. Three consecutive robust generations require consideration of another generation (eight additional cells): $F_0(R_{N=3}) =$ the total number of cells multiplied by $\langle X \rangle^{2+4+8}$ for $N = 3$.



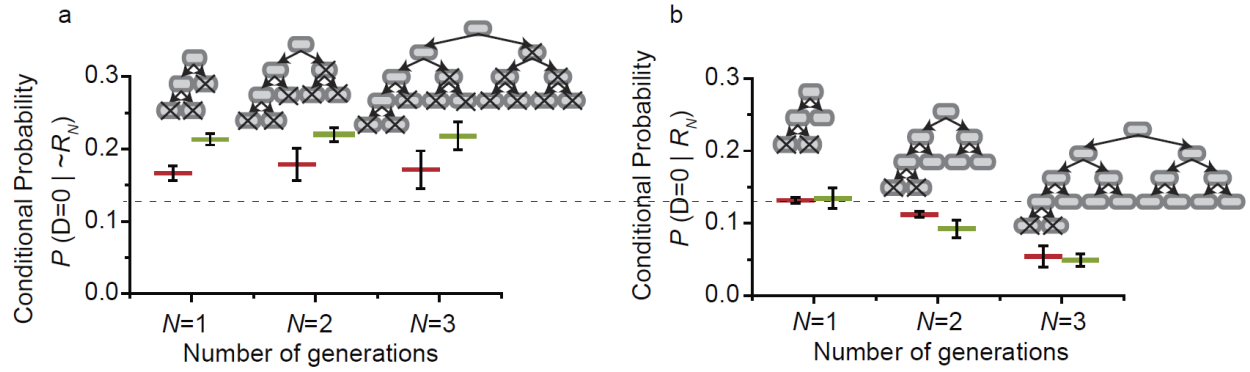
Supplementary Fig. 3. Conditional probability of one daughter cell survival ($D = 1$) given N non-robust generations (left) and robust generations (right). We analyzed how likely robust and non-robust generations give birth to one daughter cell. The left panel shows the conditional probability of one daughter cell surviving from non-robust generations. The right panel shows the conditional probability of one daughter cell surviving from robust generations for comparison. Sample sizes, number of replicates and symbols are described in detail in Fig. 2 and 3 captions.



Supplementary Fig. 4. Conditional probability of two daughter cell survival ($D = 2$) given N non-robust generations (left) and robust generations (right, Fig. 3b replot). In Fig. 3b, we plotted the conditional probability of two daughter cells from robust generations (replotted as (b) here). Here, we plotted the conditional probability of two daughter cells from non-robust generations (a). To statistically test whether $P(D = 2 | R_N)$ is significantly greater than $P(D = 2 | \sim R_N)$ for each N , we performed a one-sided t-test between them using the weighted means and standard errors (as plotted here, whose values can be found in Source data 1). The p-values are provided below. The p-values were lower than 0.05, except for two cases (WT $N=3$ and $\Delta tolC$ $N=1$). But even in these two cases, the p-values were lower than 0.08. Therefore, their differences are generally significant.

	WT	$\Delta tolC$
$N = 1$	0.033462	0.05607
$N = 2$	0.022786	0.00976
$N = 3$	0.070639	0.002373

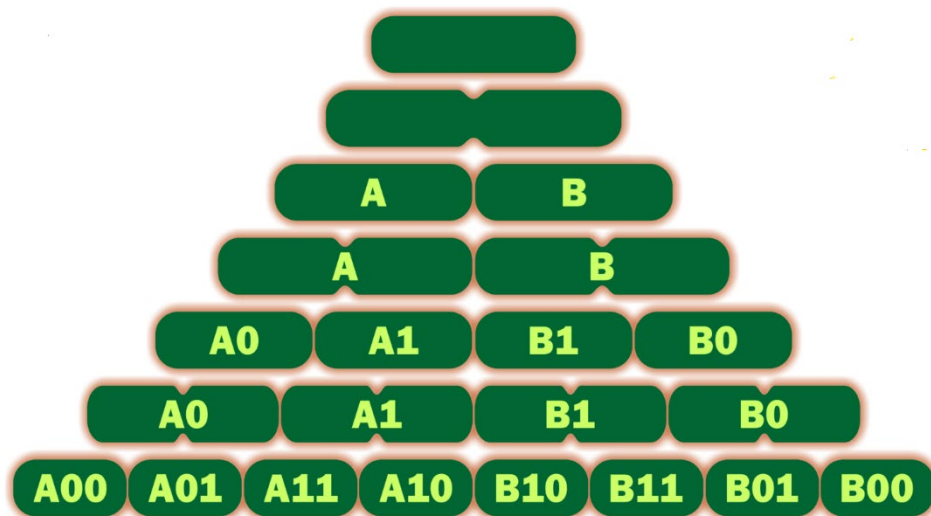
Additionally, $P(D = 2 | \sim R_N)$ changed little with N . The dashed line indicates the value of $P(D = 2 | R_N)$ for $N=1$ and is drawn for comparison with other values. Sample sizes, number of replicates and symbols are described in detail in Fig. 2 and 3 captions.



Supplementary Fig. 5. Conditional probability of zero daughter cell survival ($D = 0$) given N non-robust generations (left) and robust generations (right, Fig. 3c replot). In Fig. 3c, we plotted the conditional probability of zero daughter cells from robust generations (replotted as (b) here). Here, we plotted the conditional probability of zero daughter cells from non-robust generations (a). To statistically test whether $P(D = 0 | \sim R_N)$ is significantly greater than $P(D = 0 | R_N)$ for each N , we performed a one-sided t-test between them using the weighted means and standard errors (as plotted here, whose values can be found in Source data 1). The p-values are provided below, which show that their difference is significant.

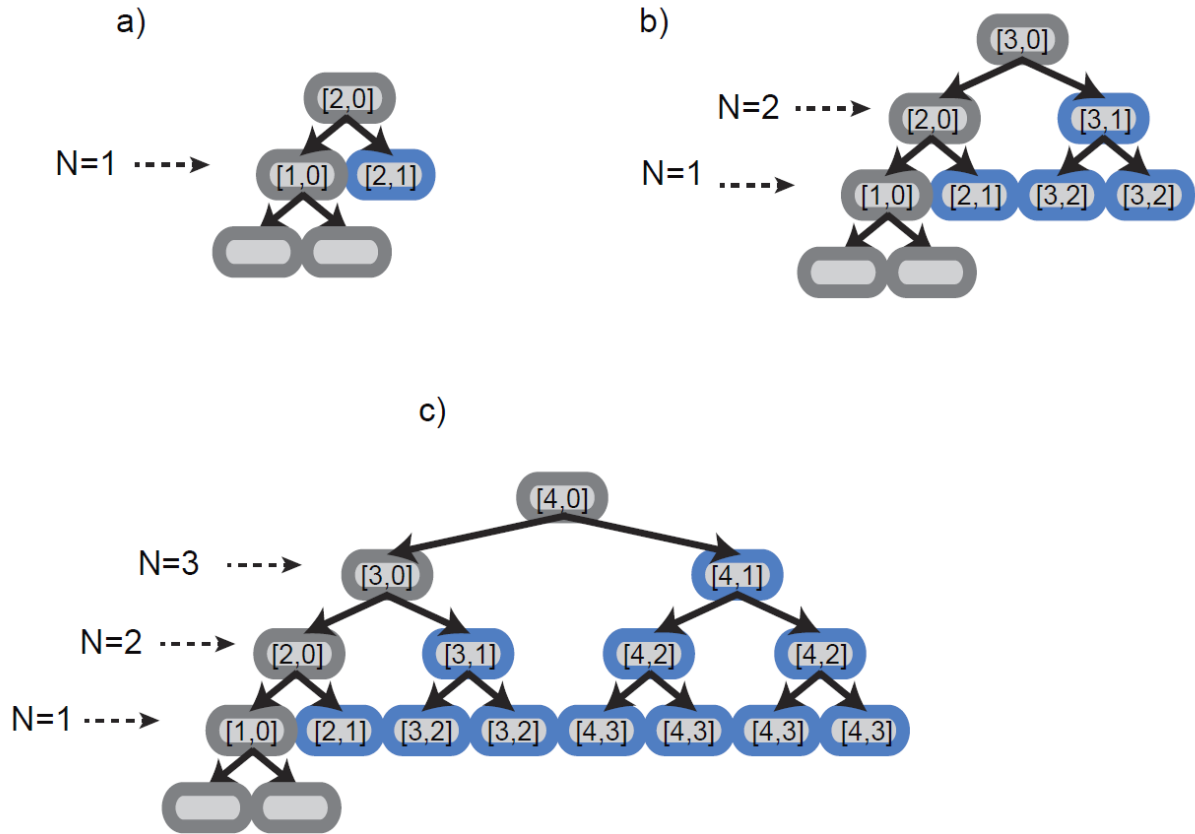
	WT	$\Delta tolC$
N = 1	0.042763	0.020048
N = 2	0.049382	0.00716
N = 3	0.029611	0.007451

Additionally, $P(D = 0 | \sim R_N)$ changed little with N . The dashed line indicates the value of $P(D = 0 | R_N)$ for $N=1$ and is drawn for comparison with other values. Sample sizes, number of replicates and symbols are described in detail in Fig. 2 and 3 captions.



Supplementary Fig. 6. Cell naming scheme.

When cell growth was analyzed using TrackMate, an ImageJ plugin, cells needed to be given unique names that would specify their place in the lineage tree, and their relationship to all other cells. When a colony's progenitor cell divides, its offspring are called "A" and "B" (since we don't know which one has the old pole). When all other cells divide, a "0" is appended to make the name of the old-pole daughter, and a "1" is appended to make the name of the new-pole daughter. At a later stage of analysis, these binary names are converted to integers as described in Methods.



Supplementary Fig. 7. Calculation of conditional Shannon entropy.

In the main text, we calculated Shannon entropy conditioned upon the fates of cells in the prior generation (see Methods). Denoting the cell fate in the prior generation by Y , we have $H(D|Y) \equiv \sum p_y H(D|Y = y)$. Here, the summation runs for all different fates of Y . Here, we described this calculation in more detail.

a). The ‘1’ generations ($N = 1$) include a mother $[1,0]$ or an aunt $[2,1]$. However, a mother’s state does not contribute to the conditional Shannon entropy because for a cell should be alive, its mother cell should always have lived. Therefore, for the conditional Shannon entropy for $N=1$, we only consider the condition of an aunt (marked as blue in Supplementary Fig. 7a). For a given cell, there is only one aunt. This aunt can live or die, which we mathematically write as $y_{[2,1]} = 1$ or 0, respectively. Then

$$H(D|Y) = \sum_{y_{[2,1]}=0,1} p_{Y=y_{[2,1]}} H(D|Y = y_{[2,1]}) \quad \text{Eq. 40}$$

b) For $N=2$, we consider not only the fate of the aunt [2,1], but also of the great aunt [3, 1] and the first cousin once removed [3, 2] (marked as blue in Supplementary Fig. 7b). We denote the combination of their fates by $\{y_{[2,1]}, y_{[3,1]}, y_{[3,2]}\}$. Similar to an aunt [2,1], there is only one cell in the relationship defined by [3,1]. Therefore, $y_{[2,1]}=1$ or 0 and $y_{[3,1]}=1$ or 0. However, there are two cells in the relationship defined by [3,2]; see Supplementary Fig. 7b. Therefore, $y_{[3,2]}$ runs from 2 (both cells live), 1 (only one cell lives), and 0 (both no cells lived); thus $y_{[3,2]}=2, 1$, or 0.

$$H(D|Y) = \sum_{\substack{y_{[2,1]}=0,1, \\ y_{[3,1]}=0,1, \\ y_{[3,2]}=0,1,2}} p_{Y=\{y_{[2,1]}, y_{[3,1]}, y_{[3,2]}\}} H(D|Y = \{y_{[2,1]}, y_{[3,1]}, y_{[3,2]}\}) \quad \text{Eq. 41}$$

c) For $N=3$, we consider the fates of the cells with the relationship [2, 1], [3, 1], [3, 2], [4, 1], [4, 2] and [4, 3]; marked as blue in Supplementary Fig. 7c. [3,3] is not included here because it is in the same generation as a self-cell. There is only one cell in the relationship defined by [4,1]; Supplementary Fig. 7c. Thus, $y_{[4,1]}=1$ or 0. There are two cells in the relationship defined by [4,2], while four cells for [4, 3]. Therefore, similar to our description above, $y_{[4,2]}=2, 1$, or 0, and $y_{[4,3]}=4, 3, 2, 1$, or 0. Hence,

$$H(D|Y) = \sum_{\substack{y_{[2,1]}=0,1, \\ y_{[3,1]}=0,1, \\ y_{[3,2]}=0,1,2, \\ y_{[4,1]}=0,1, \\ y_{[4,2]}=0,1,2, \\ y_{[4,3]}=0,1,2,3,4}} p_{Y=\{y_{[2,1]}, y_{[3,1]}, y_{[3,2]}, y_{[4,1]}, y_{[4,2]}, y_{[4,3]}\}} H(D|Y = \{y_{[2,1]}, y_{[3,1]}, y_{[3,2]}, y_{[4,1]}, y_{[4,2]}, y_{[4,3]}\}), \quad \text{Eq. 42}$$

Supplementary reference:

- 1 Kot, M. *Elements of Mathematical Ecology*. (Cambridge University Press, 2001).
- 2 Coates, J. *et al.* Antibiotic-induced population fluctuations and stochastic clearance of bacteria. *eLife* **7**, e32976 (2018). <https://doi.org/10.7554/eLife.32976>