Supplemental Information

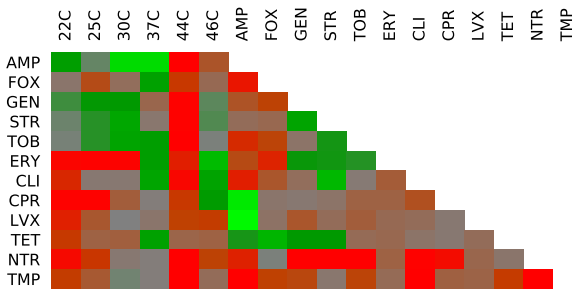
Interactions and clusters through time

The distribution of the measured interactions between the stressors changes when growth is measured at different timepoints (Supplemental Figure 2). Most calculated interactions are antagonistic at early timepoints (4, 8 hr), and using the clustering procedure in these timepoints fails to separate the antibiotics by their mechanism of action (Supplemental Figure 3), since there are not enough differences in the measured interactions between the stressors to clearly distinguish the clusters. At later timepoints (12h, 24h), there is a gradual shift towards more synergistic interactions, and the clustering successfully separates the stressors into functionally related classes.

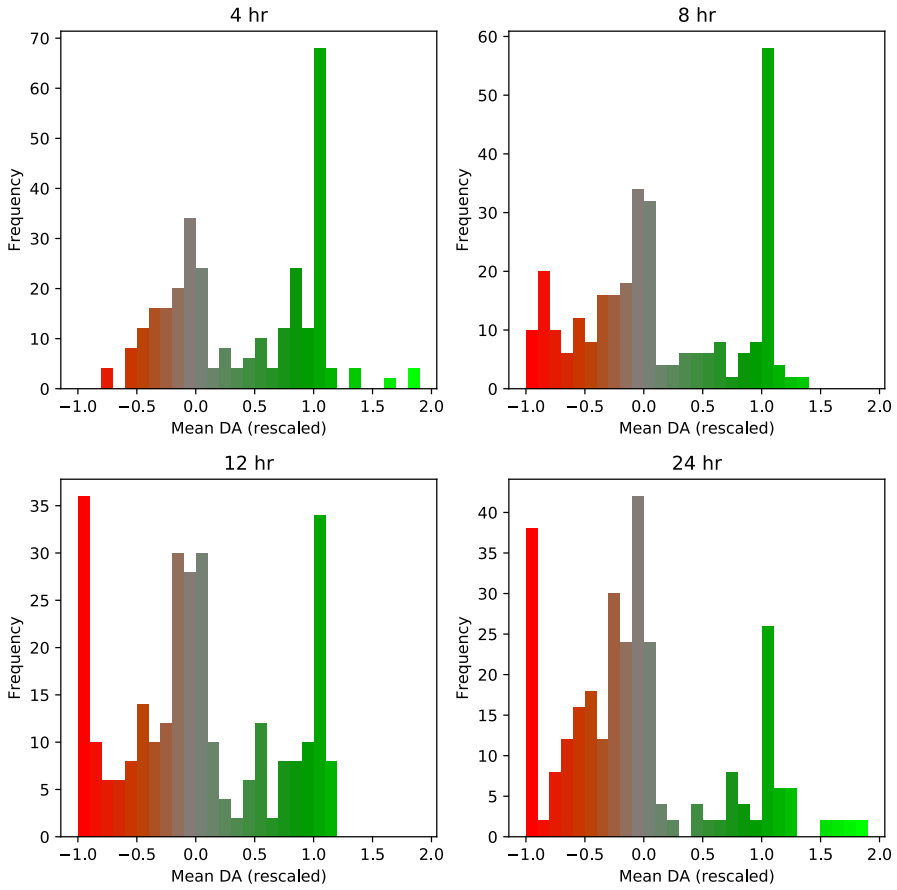
We attribute the lack of information to distinguish the clusters in early timepoints to two factors: 1) The interactions at early timepoints depend on detecting smaller differences in growth across the different conditions than in later timepoints. Furthermore, the interactions must be inferred from measurements of low optical densities, which have large measurement errors. Because of this, the measured interactions between stressors are much noisier and harder to distinguish from each other. After the bacterial populations have had more time to grow under each condition, differential growth patterns are easier to discern and the OD values are more reliable. 2) Some effects of the antibiotics and temperatures in cell physiology (which include the cellular response) may not be fully realized until later timepoints.

At 12h, monochromatic clustering successfully separates the antibiotics into their mechanisms of action, but the resulting clustering still shows some inconsistencies (such as grouping the lowest, middle and highest temperatures together). Because of this, we chose to use the 24h timepoint for the analysis. At this timepoint, all antibiotics which share the same mechanism of action form part of the same cluster and the temperatures cluster in a consistent way (with no clusters that have skipped intermediate temperatures).

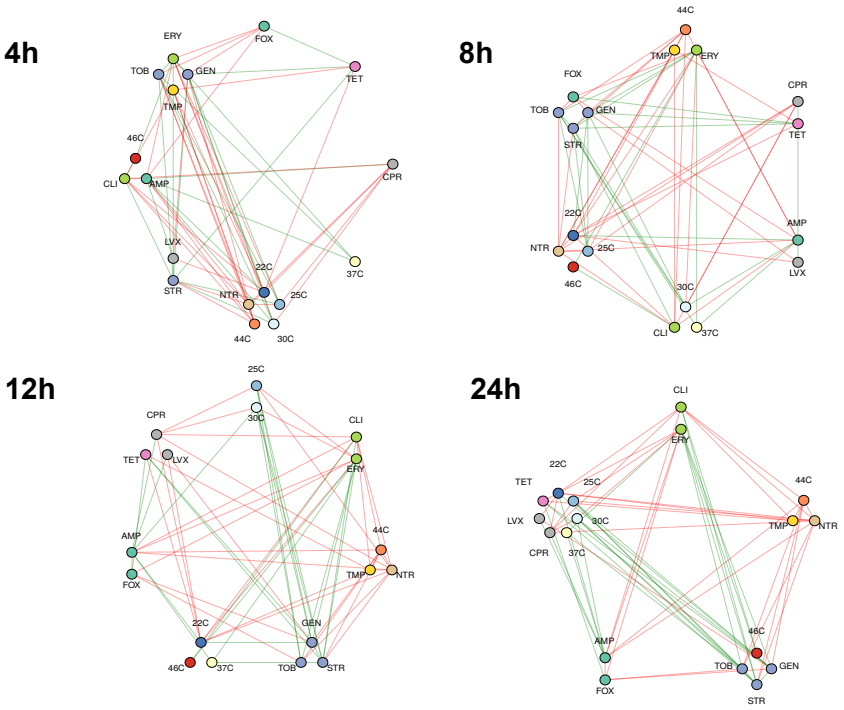
Supplemental Figures



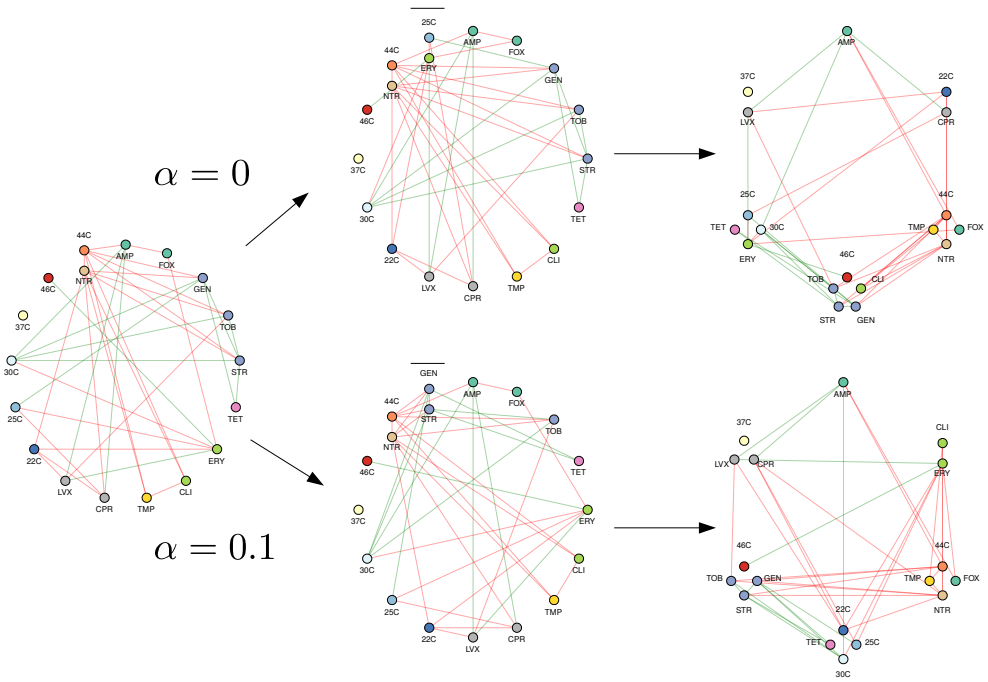
Supplemental Figure 1: *Median interaction effects between antibiotics and temperature after 24 hour growth.* The interaction effect ($\tilde{\varepsilon}$) values are color coded in a gradient, from synergy (red) to additive (grey) and antagonism (green).



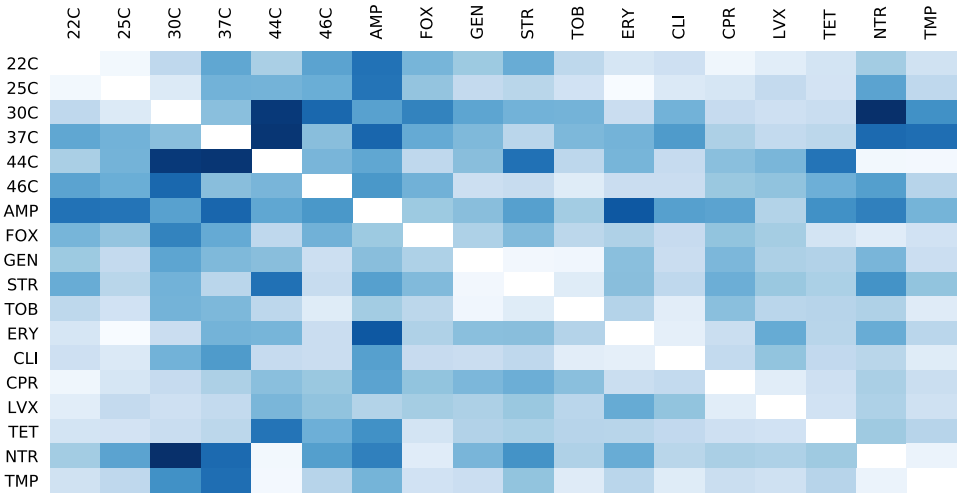
Supplemental Figure 2: *Time-resolved distribution of interaction effects.* The distribution of the mean antibiotic and temperature interaction effects at the four measured timepoints (4, 8, 12 and 24 hr) is shown. At hour 4, synergistic interactions are rare. As time goes by, the antibiotics gradually show their full effects on growth. This results in more synergistic effects being apparent, and an overall shift in the distribution to the left. By hour 24, four modes are clearly visible in the distribution, corresponding to strong antagonism (≈ -1), additivity (≈ 0), and moderate (≈ 0.5) and strong (≈ 1) antagonistic buffering.



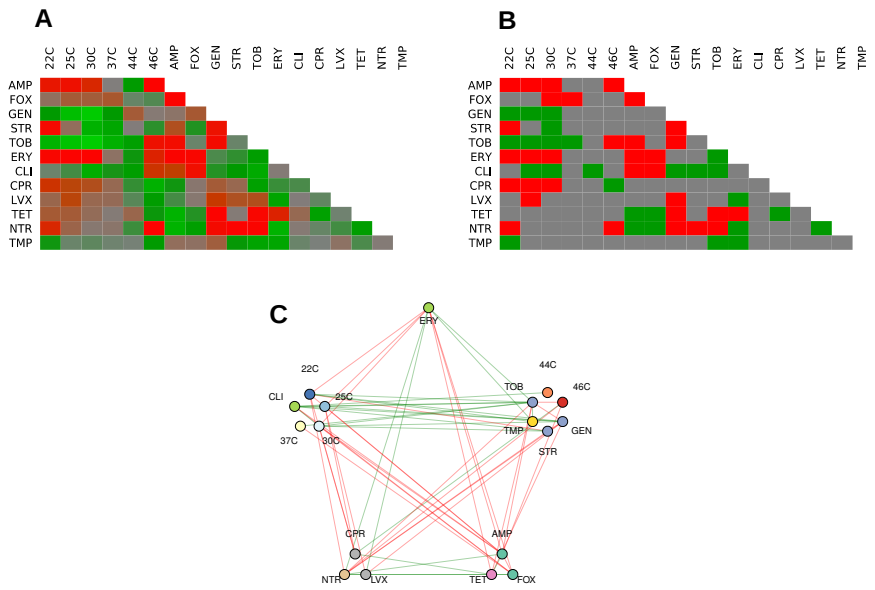
Supplemental Figure 3: *Time-resolved interaction networks.* Clusterings constructed based on early time point measurements of growth are uninformative because the observed interactions are less reliable (due to their dependence on measurements of low optical densities (OD)) and the effect of the antibiotics not being fully realized until later time points.



Supplemental Figure 4: *Effect of the non-shared interaction penalty α in the clustering algorithm.* The clusters obtained when discretizing the interactions as in [**?**] (producing fewer edges in the network than the method used in the main text) is shown for $\alpha = 0$ and $\alpha = 0.1$. The penalty prevents the merging of clusters with no or few shared edges, thus improving the quality of the clustering as evaluated by cluster membership of antibiotics with the same mechanism of action. This penalty makes an especially large difference in networks that have few edges.



Supplemental Figure 5: *Dissimilarity of interactions.* A heatmap with the dissimilarity between the interactions of each pair of conditions (see Methods) is shown. The dissimilarities are color coded in a gradient from white (more similar) to dark blue (more dissimilar). From these results, it is apparent that temperatures can be similar to multiple classes of antibiotics.



Supplemental Figure 6: *Interactions and clustering under no salt conditions.* The interaction effect ($\tilde{\varepsilon}$) values are color coded in a gradient, from synergy (red) to additive (grey) and antagonism (green). (a) Matrix heatmap of the mean interaction effects. Antibiotics with the same mechanism of action show similar interaction patterns. (b) Matrix heatmap of the discretized interaction types used for constructing the edges of the interaction network. (c) Network clustered into monochromatic classes by the modified Prism2 algorithm (see Methods).

Species	Background	Notes	Strain
E. coli	REL1206	High-temperature adapted line	HA1
E. coli	REL1206	High-temperature adapted line	HA2
E. coli	REL1206	High-temperature adapted line	HA3
E. coli	REL1206	High-temperature adapted line	HA4
E. coli	REL1206	High-temperature adapted line	HA5
E. coli	REL1206	High-temperature adapted line	HA6
E. coli	REL1206	High-temperature adapted line	HA7
E. coli	REL1206	High-temperature adapted line	HA8
E. coli	REL1206	High-temperature adapted line	HA9
E. coli	REL1206	High-temperature adapted line	HA10
E. coli	REL1206	Ancestor	
E. coli	REL1206	REL1206 with mutation rpoB I572F	M1
E. coli	REL1206	REL1206 with mutation rpoB I572L	M2
E. coli	REL1206	REL1206 with mutation rpoB I572N	M3

Supplemental Table 1: Information of temperature-adapted and related strains.

Supplemental methods

Rescaling of drug and temperature interactions

The raw interaction values $\varepsilon_{xy} = w_{xy} - w_x w_y$ in the same way as in [1], in accordance to the table below.

Interaction type	Growth rates condition	Rescaled interaction ($\tilde{\varepsilon}_{xy}$)	$\tilde{\varepsilon}_{xy}$ interval
Synergy	$w_{xy} < w_x w_y$	$\frac{\varepsilon_{xy}}{\lvert 0 - w_x w_y \rvert}$	$[-1, 0)$
Antagonistic buffering	$w_x w_y \leq w_{xy} < \min(w_x, w_y)$	$\frac{\varepsilon_{xy}}{\lvert \min(w_x, w_y) - w_x w_y \rvert}$	$[0, 1)$
Antagonistic suppression	$w_{xy} \geq \min(w_x, w_y)$	$1 + \frac{\varepsilon_{xy}}{\lvert 1 - w_x w_y \rvert}$	$[1, \infty)^*$

Supplemental Table 2: Definition of the rescaled deviation from additivity $\tilde{\varepsilon}_{xy}$ for the synergy, antagonistic buffering and suppression cases, and the interval of values $\tilde{\varepsilon}_{xy}$ can take for each interaction type. * While mathematically possible, $\tilde{\varepsilon}_{xy}$ values greater than two are uncommon, since it is almost always true that $w_{xy} \leq 1$ (i.e. the growth under conditions (x, y) simultaneously is less than the growth under optimal conditions).

The modified Prism2 algorithm for monochromatic clustering

The modified Prism2 algorithm used in this work is a variant of hierarchical clustering, with an added entropy term to penalize non-monochromaticity. Each node starts in a different cluster, and at each iteration of the algorithm the pair of clusters (X, Y) that minimizes the penalty

$$F(X, Y) = k_D D(X, Y) + k_S \left(\Delta S(X, Y) + \alpha(1 - p_{XY}) \right)$$

is merged. This procedure is repeated until a single cluster remains, containing all nodes in the network.

The first term in $F(X, Y)$ corresponds to the standard cost term in hierarchical clustering. It penalizes merging clusters that are dissimilar to each other in terms of their interactions with other clusters. The interaction dissimilarity between nodes x, y is defined as

$$d(x, y) = \frac{1}{N(x, y)} \sum_{z \neq x, y} \left(\frac{\hat{\tilde{\varepsilon}}_{xz} - \hat{\tilde{\varepsilon}}_{yz}}{2} \right)^2$$

where $N(x, y)$ is the number of interactions with other nodes z≠x,y (i.e., number of $(\tilde{\varepsilon}_{xz}, \tilde{\varepsilon}_{yz})$ pairs) that were measured for both conditions x and y. The mean value of the dissimilarity between all pairs of nodes belonging to clusters (X, Y)

$$D(X, Y) = \frac{1}{n_X \cdot n_Y} \sum_{x \in X, y \in Y} d(x, y)$$

where n_X, n_y are the number of nodes in clusters X and Y, respectively, is taken as a measure of the dissimilarity between the clusters. This is another difference

with respect to the previous algorithm. The use of the mean distance between the antibiotics (average linkage) as the distance between clusters $D(X, Y)$ is less sensitive to outliers than the previous choice of the minimum distance (single-linkage).

The second term in the penalty is based on the information-theoretic concept of entropy, and penalizes non-monochromatic interactions between clusters X and Y. The interaction entropy between two clusters is defined as

$$S(\vec{m}_{X,Y}) = -(m^+_{X,Y} + m^-_{X,Y})(p^+_{X,Y} \log_2 p^+_{X,Y} + p^-_{X,Y} \log_2 p^-_{X,Y})$$

where $m^+_{X,Y}$ and $m^-_{X,Y}$ are the number of red (synergy) and green (antagonism) edges between clusters X and Y, and $p^+_{X,Y} = \frac{m^+_{XY}}{m^+_{XY}+m^-_{XY}}$, $p^-_{X,Y} = \frac{m^-_{XY}}{m^+_{XY}+m^-_{XY}}$ are the corresponding proportions (we use the vector $\vec{m}_{X,Y} = (m^+_{XY}, m^-_{XY})$ as a shorthand for notational simplicity). Only the entropy of between-clusters interactions is penalized (but not the entropy of interactions where both nodes are part of the same cluster). When a pair of clusters X, Y is merged into cluster $M = X \cup Y$, the entropy of their interactions $S(\vec{m}_{X,Y})$ is "hidden" inside the new cluster and no longer penalized. The change in the interaction entropy across all the network upon merging is

$$\begin{aligned}
\Delta S(X,Y) &= S_{\text{gained}} - S_{\text{lost}} \\
&= \sum_{Z \neq X,Y} [S(\vec{m}_{M,Z}) - S(\vec{m}_{X,Z}) - S(\vec{m}_{Y,Z})] - S(\vec{m}_{X,Y}) \\
&= \sum_{Z \neq X,Y} [S(\vec{m}_{X,Z} + \vec{m}_{Y,Z}) - S(\vec{m}_{X,Z}) - S(\vec{m}_{Y,Z})] - S(\vec{m}_{X,Y})
\end{aligned}$$

where S_{gained} is the net change in entropy over all interactions X and Y had with all other clusters, and $S_{lost} = S(\vec{m}_{X,Y})$ is the entropy lost by the newly "hidden" interactions between X and Y. Note that in this formulation we chose to write ΔS with an opposite sign as in [1], to be consistent with the usual convention that positive values mean increases and negative values decreases.

A third ingredient in the algorithm, which is not present in previous versions, is the term $\alpha(1 - p_{XY})$, where p_{XY} is the proportion of shared edges between clusters X, Y and α an arbitrary tuning constant. It is always possible to form monochromatic clusters by simply merging clusters with members which have no, or few, shared edges, but these clusters need not have similar overall interactions (and thus lack any physiological relevance). This is particularly important in the earlier steps of the algorithm, where most clusters are small, so there is a larger proportion of clusters with missing or unknown edge colors. The purpose of this term is to avoid this problem by penalizing the entropy term so joining clusters with more shared interactions is favored with respect to clusters with few (Supplemental Figure 4).

In the penalty $F(X, Y)$, each term is multiplied by constants that affect the relative weight of each term. We chose $k_D = 1$, $k_S = 0.1$ to be the same values

as in [1]. The third tuning constant, $\alpha = 0.2$, is new to this work, and was chosen to be of roughly the same order of magnitude, but smaller, than the entropy $\Delta S_{XY} \in [0, 1]$.

Promoter library data analysis

Background removal

The promoter library is split into six 384-well plates. Besides the GFP-expressing strains of the library, which are fused to different *E. coli* promoters, the plates also contain empty wells and strains with a promoterless plasmid as controls. The raw measured OD values were background-corrected by subtracting the mean OD of the empty wells in the same plate at each timepoint. Since the measurements were done in the absence of the resistance marker of the plasmids (KAN) to avoid interference with the response to the antibiotics, empty wells would ocassionally become contaminated with a neighboring strain. Because of this, the data was first processed to remove contaminated wells before calculating the mean. This was done by excluding all empty wells for which the raw measured OD was more than 0.3 by the last measured timepoint (20 hours after exposure to the stressor) from the background calculation.

Background subtraction of the raw GFP fluorescence measurements was done relative to the promoterless strains in the corresponding plate. A simple strategy of substracting the mean fluorescence of the promoterless strain is unsatisfactory since a small number of genes with small fluorescence yield negative values, which prevent the calculation of log-fold changes. Because of this, we followed a background-removal strategy similar to that used in RMA, a popular method for normalizing microarray data [2, 3]. This is a model-based btechnique in which the measured raw fluorescence intensity X_i is assumed to be a sum of normally distributed background B_i and strictly positive, exponentially distributed, signal S_i. That is, the raw fluorescence measurements are modeled as

$$\begin{aligned} X_i &= B_i + S_i \\ B_i &\sim N(\mu, \sigma^2) \\ S_i &\sim \text{Exponential}(\alpha) \end{aligned}$$

An estimate of the signal (i.e. background-removed measurement) is obtained by calculating the expected value of the signal given the measured intensity X_i. The parameters were estimated with the "nonparametric estimator" detailed in [3]. Briefly, the parameters (μ, σ^2) of the background were estimated with the sample mean and variance of the promoterless wells (which are assumed to have only background, no signal) in the corresponding plate. The parameter α corresponds to the mean of the signal. It was estimated with the difference between the sample means of the measurements of the strains with promoters and the promoterless strains. Once the parameters are estimated, a closed form solution, also available in [3], was used to calculate the background-removed measurement for all promoter-containing strains.

Data processing

Since the experimental conditions resulted in different overall growth, the background-removed GFP fluorescence measurements at each timepoint were normalized by the corresponding OD mesaurement to get a quantity proportional to the amount of fluorescence per cell. To make replicates comparable, the resulting GFP/OD values were median-normalized. The median-normalized values were used to calculate log2-fold changes with respect to the control condition (37°C). Four replicates of the log2-fold change were averaged to yield a final value of the log2-fold change in expression at each timepoint of each experimental condition (44°C, STR, TET) compared to control.

Finding over-expressed and under-expressed genes

A set of overexpressed and underexpressed genes under each experimental condition was found by the robust z-score method, as in (Reyes et al., 2012). The robust z-score for promoter p in condition c and time t is given by

$$z_{cpt}^{(r)} = \frac{x_{cpt} - \text{median}(x_{ct})}{\text{MAD}(x_{ct})}$$

where x_{cpt} is the log2-fold change with respect to control of the corresponding promoter, condition and time, $\text{median}(x_{ct})$ is the sample median of the log2-fold change measurements for all promoters at that timepoint and condition, and

$$\text{MAD}(x_{ct}) = 1.4826 \times \text{median}\left(|x_{cpt} - \text{median}(x_{ct})|\right)$$

is the median absolute deviation. A gene was considered overexpressed at time t if $z_{cpt}^{(r)} > 2.5$ and underexpressed if $z_{cpt}^{(r)} < -2.5$. The genes overexpressed and underexpressed at all timepoints were pooled together to yield a single set of overexpressed and underexpressed genes for each condition.

Finding IC50 values

The five parameter logistic model

IC50 values were found by fitting a five-parameter logistic model, which allows for assymetric dose-response curves, to the drug concentration vs growth data for each antibiotic. The model is given by

$$g(c) = g_{\min} + (g_{\max} - g_{\min}) \left(\frac{1}{1 + \left(\frac{c_b}{c}\right)^n} \right)^s$$

where $g(c)$ is the growth as a function of the drug concentration c. The IC50 can be calculated as

$$\text{IC50} = \frac{c_b}{(2^{\frac{1}{s}} - 1)^{\frac{1}{n}}}$$

Inference of model parameters

This model was fit with a Bayesian procedure, by extending it to a statistical model. It is important to make the distinction between the observed data values and the underlying parameters of the model. In statistics, model parameters are typically given Greek letter notation and observed values Roman letters, but this convention is inconvenient when dealing with deterministic models, which use Roman letters throughout. To avoid confusion, we will use the same Roman letters for parameters that come from the deterministic model, and only use Greek letters for extra parameters that are needed for the statistical model (e.g. variances).

Let $\mathcal{P} = \{g_{\min}, g_{\max}, c_b, n, s\}$ be the parameters of the logistic model and y_{ci} be the i-th replicate of the OD measurement at drug concentration c. We fit a model of the form

$$\begin{aligned} y_{ci}|\mathcal{P}, \sigma &\sim \text{Gamma}(\mu = g(c),\, \sigma = \sigma_y(c)) \\ \sigma_y(c) &= \sigma_{\min} + \beta\,[g(c) - g_{\min}] \end{aligned}$$

The model for the standard deviation was motivated by the observation that the growth measurements tend to have low variance when there is essentially no growth (at high concentrations), and that the variance increases as the OD increases. A linear form was chosen for simplicity.

We used the following priors:

$$\begin{aligned} \text{IC50} &\sim \text{Uniform}(0,\, 500) \\ \Delta g &\sim \text{Uniform}(0,\, 1) \\ g_{\min} &\sim \text{Gamma}(\mu = 0.05, \sigma = 0.02) \\ \ln n &\sim \text{Normal}(0,\, 1.5) \\ \ln s &\sim \text{Normal}(0,\, 1.5) \\ \sigma_{\min} &\sim \text{Gamma}(\mu{=}0.02, \sigma = 0.02) \\ \beta &\sim \text{Gamma}(\mu = 0.1, \sigma = 0.3) \end{aligned}$$

(where we reparametrized the model). To recover the original parameters $\mathcal{P}$, we have that

$$\begin{aligned} g_{\max} &= g_{\min} + \Delta g \\ n &= \exp(\ln n) \\ s &= \exp(\ln s) \\ c_b &= \text{IC50}(2^{\frac{1}{s}} - 1)^{\frac{1}{n}} \end{aligned}$$

The motivation for the non-uniform priors is as follows. In the five-parameter logistic model, n determines the steepness of the change in growth as a function of concentration, and s allows for assymetric curves. When $s = 1$, the curve is symmetric and the model reduces to a Hill equation. When $s = 1$ and $n = 1$, the model reduces to a simple Michaelis-Menten-like form. We chose to

use weakly informative priors for n and s as a form of regularization to keep the values for n and s in a reasonable range (roughly $\frac{1}{20}$ to 20). We chose to give greater prior probability to models where n and s are close to one due to the connection to simpler models. The informative prior for $\sigma_{\min}$ represents that we expect bacterial growth to be essentially zero at very high antibiotic concentrations, and the variance to consist mostly of measurement error for OD, which is typically below 0.02 in OD units.

Model fitting

For each antibiotic and strain, the model above was fit to the data using NUTS (the no-u-turn sampler), a variant of Hamiltonian Monte Carlo, as implemented in the Python library PyMC3 [5]. Three chains were run for 5000 iterations, after an initial 1500 samples were used for tuning. The output samples from the posterior distribution were used to construct a mean estimate and a credible interval for the IC50 for each antibiotic and strain. For a few strain-antibiotic combinations, the NUTS chain did not converge (as evaluated by $\hat{R} > 1.2$) or had a low number of effective samples ($n_{\text{eff}} < 500$). These estimates were considered unreliable and were removed from the plots and analysis.

References

[1] Yeh, P., Tschumi, A. I., & Kishony, R. (2006). Functional classification of drugs by properties of their pairwise interactions. Nature genetics, 38(4), 489.

[2] Irizarry, R. A., Wu, Z., & Jaffee, H. A. (2006). Comparison of Affymetrix GeneChip expression measures. Bioinformatics, 22(7), 789-794.

[3] Xie, Y., Wang, X., & Story, M. (2009). Statistical methods of background correction for Illumina BeadArray data. Bioinformatics, 25(6), 751-757.

[4] Reyes, V. C., Li, M., Hoek, E. M., Mahendra, S., & Damoiseaux, R. (2012). Genome-wide assessment in Escherichia coli reveals time-dependent nanotoxicity paradigms. ACS nano, 6(11), 9402-9415.

[5] Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. PeerJ Computer Science, 2, e55.