

RESEARCH ARTICLE

# Parallel Mapping of Antibiotic Resistance Alleles in *Escherichia coli*

Sophie J. Weiss<sup>1</sup>✉, Thomas J. Mansell<sup>1</sup>✉, Pooneh Mortazavi<sup>2</sup>, Rob Knight<sup>3,4</sup>, Ryan T. Gill<sup>1\*</sup>

**1** Department of Chemical and Biological Engineering, University of Colorado Boulder, 3415 Colorado Avenue, Boulder, Colorado, 80303, United States of America, **2** Department of Computer Science, University of Colorado Boulder, 1111 Engineering Drive ECOT 717, Boulder, CO 80303, United States of America, **3** Department of Pediatrics, University of California San Diego School of Medicine, 9500 Gilman Drive, MC 0602, La Jolla, CA 92093, United States of America, **4** Department of Computer Science & Engineering, University of California San Diego, 9500 Gilman Drive, MC 0404, La Jolla, CA 92093, United States of America

✉ These authors contributed equally to this work.

\* [rtg@colorado.edu](mailto:rtg@colorado.edu)



## Abstract

Chemical genomics expands our understanding of microbial tolerance to inhibitory chemicals, but its scope is often limited by the throughput of genome-scale library construction and genotype-phenotype mapping. Here we report a method for rapid, parallel, and deep characterization of the response to antibiotics in *Escherichia coli* using a barcoded genome-scale library, next-generation sequencing, and streamlined bioinformatics software. The method provides quantitative growth data (over 200,000 measurements) and identifies contributing antimicrobial resistance and susceptibility alleles. Using multivariate analysis, we also find that subtle differences in the population responses resonate across multiple levels of functional hierarchy. Finally, we use machine learning to identify a unique allelic and proteomic fingerprint for each antibiotic. The method can be broadly applied to tolerance for any chemical from toxic metabolites to next-generation biofuels and antibiotics.

## OPEN ACCESS

**Citation:** Weiss SJ, Mansell TJ, Mortazavi P, Knight R, Gill RT (2016) Parallel Mapping of Antibiotic Resistance Alleles in *Escherichia coli*. PLoS ONE 11 (1): e0146916. doi:10.1371/journal.pone.0146916

**Editor:** Brian Frederick Pfleger, University of Wisconsin, UNITED STATES

**Received:** May 4, 2015

**Accepted:** December 23, 2015

**Published:** January 15, 2016

**Copyright:** © 2016 Weiss et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Raw FASTQ files have been uploaded to the NCBI BioProject accession number PRJNA260093. A table of sequence counts is also available in the Supporting Information.

**Funding:** SJW was funded by the NIH/CU Molecular Biophysics Training Scholarship (T32 GM-065103).

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

Chemical genomics, or the study of the genome-scale response to small molecules, has rapidly advanced thanks to synthetic biology approaches. For example, studies of phenotype mapping of small molecule landscapes have led to elucidation of novel genetic functions and drug mechanisms [1–3]. These pioneering studies took large genomic libraries, usually painstakingly created [4, 5], and characterized them under a range of chemical and physical conditions using DNA microarrays. Studies of chemical tolerance have also used adaptive evolution methods to identify mutations that contribute to fitness [6, 7]. While these studies closely mimic responses to stresses in nature, the extent of genotyping is limited by the throughput of whole-genome sequencing.

The increasing throughput and decreasing cost of multiplex oligonucleotide synthesis [8] and high-throughput sequencing [9] has enabled unprecedented advances in throughput of

genome engineering and analysis technologies [10–13]. For example, recent studies have leveraged high-throughput sequencing to expand the characterization of yeast deletion libraries [14]. Along these lines, we recently reported the trackable multiplex recombineering (TRMR) approach [15]: a one-pot construction of a barcoded, genome-scale library simulating overexpression and knockdown of over 4,000 genes in the Gram-negative bacterium *E. coli*. Initial experiments with the library focused on the genomic response to various carbon sources and biofuel-related inhibitory conditions using DNA microarrays and exploratory by-hand analyses [15, 16].

At sub-lethal antibiotic concentrations such as those found in wastewater and agricultural runoff, the contribution to microbial fitness of cellular factors is not nearly as well-studied [17] as horizontal gene transfer of specific resistance effectors [18]. Thus, understanding the response and resistance of microbes to antimicrobial compounds is of critical importance. To isolate gene products contributing to antibiotic resistance, several genomic and proteomic studies have been performed [19–23]. However, previous attempts to characterize genome-scale responses to antibiotic challenges [1, 3, 6, 7, 24–26] relied on either (1) the low-throughput construction of large libraries or (2) many generations of adaptive evolution, where characterization was limited by sequencing surviving colonies.

Here we report a method for the rapid and deep characterization of laboratory population dynamics in response to eight antibiotics by multiplex selection, next-generation sequencing, and multivariate analysis of *E. coli* TRMR libraries. Our findings support the development of multi-drug resistance and susceptibility genes as an important step in the evolution of antibiotic resistance in microbial populations at sub-lethal concentrations. Finally, to expand the throughput and extent of our bioinformatic analysis, we integrate the data gathered into the QIIME multivariate analysis pipeline, with which we examine the response at a pathway level and identify a unique genomic signature for each antibiotic.

## Methods

### Strains and plasmids

The TRMR library was previously constructed [15]. Briefly, *E. coli* MG1655 (ATCC #700926) cells were subjected to multiplex recombineering using synthetic DNA cassettes containing either an “up” (strong promoter and RBS) or “down” (no promoter or RBS) phenotype with homology regions corresponding to over 4,000 genes in the *E. coli* genome. The synthetic cassettes also contained unique barcodes for rapid characterization and gene-trait mapping. In this study, a modified version of strain JWKAN, which is MG1655 with the kanamycin resistance gene *neoR* (from pKD13 [27]) inserted in a safe region and barcoded as in the rest of the library, was used as the wild-type control. Expression of FLP recombinase (pCP20 [28]) excised *neoR* from the genome using flanking FRT sites to create a barcoded MG1655 without kanamycin resistance, which we refer to as MG1655-BC. This phenotype was confirmed by replica plating and the genotype confirmed by colony PCR.

### Antibiotic MIC Determination

Overnight cultures of MG1655-BC cells were subcultured into various concentrations of antibiotics in MOPS media [29] at 37°C to determine the minimum inhibitory concentration (MIC) for each compound. All antibiotics were purchased from Sigma-Aldrich (St. Louis, MO). The MIC for each antibiotic was determined by an iterative process using the procedures and definitions of Andrews [30]. First, an estimate was determined by growing MG1655-BC in 96-well plates in triplicate in 2-fold increments around the MIC found in the literature (if any) [30]. The 2-fold determined MIC was then refined by growth in 1.2 fold increments. The

refined MIC was used for liquid culture in MOPS media in 250 mL flasks, inoculated at  $OD_{600}$  0.02 with MG1655-BC or the recovered TRMR library. The final MIC concentration was determined to be the concentration at which MG1655-BC showed no growth and the TRMR library showed significant ( $OD_{600} > 0.2$ ) growth at 24 hours.

### Cell culture and selection conditions

The TRMR “up” and “down” libraries were recovered from frozen stocks by inoculating glycerol stocks of the constructed libraries in low salt LB media with 90 $\mu$ g/mL blasticidin-S to  $OD_{600}$  0.4. The cells were grown at 37C in a shaking incubator to an  $OD_{600}$  of approximately 0.8.

When the initial TRMR and MG1655 cultures reached the desired  $OD_{600}$ , they were transferred to two identical sets of three selection flasks containing 50 mL MOPS media at 80% of the previously determined MIC (sub-inhibitory selection concentration, SSC) for each of the eight antibiotics (for 48 flasks total) tested to an  $OD_{600}$  of approximately 0.02 [15]. TRMR “up” and TRMR “down” libraries were added in equal amounts as determined by OD. These initial cultures were then harvested by centrifugation and frozen as pellets for initial concentration values, which we refer to as time point zero. Growth proceeded under antibiotic selection conditions at 37°C and cells were harvested by centrifugation after 24 hours and upon reaching a 1.5  $OD_{600}$ .

### Antibiotic Colony Sequencing

Individual colonies from each selection were amplified including the barcode tags by PCR. All PCRs were performed using Phusion polymerase (NEB). The PCR product was confirmed to correspond to the barcode region by gel electrophoresis. The DNA was then purified using a QIAquick gel extraction kit (Qiagen), and sent for Sanger sequencing (MWG Eurofins Operon). The incorporated tag sequence was compared with [S1 Table](#) of Warner et al. [15] to identify alleles. For high-throughput sequencing, the genomic DNA from  $10^9$  cells from all the selections was extracted using the DNeasy Blood & Tissue Kit (Qiagen). Four base-pair tags were appended using PCR near the beginning of each TRMR-unique barcode to further distinguish the samples by replicate. PCR products of roughly 180 bp were gel-extracted and purified using the QIAquick gel extraction kit (Qiagen), and combined in equimolar amounts. The resulting mixture of amplicons from all replicates and time points for each antibiotic sample was assigned a unique Illumina index and prepared for sequencing according to Illumina TruSeq 1x50 guidelines [31] and sequenced on an Illumina HiSeq 2000.

### Sequencing Data Analysis

Each FASTQ file produced by the high-throughput sequencing was read and signal quality filtered in parallel using a custom MATLAB script. The 50 base pair reads were matched to 50 base pair DNA sequences in a mapping file corresponding to the expected barcodes in genomic context. These sequences included a four base pair tag for replicate and experiment identification, as well as the unique TRMR tag sequences for each gene as found in [S1 Table](#) of Warner et al [15]. Any FASTQ sequence not matching those in the mapping file within 1 base pair was discarded to allow distinguishing between the replicates while minimizing spurious mapping of sequences to genes. This strict quality filtering meant 10–40% of the sequences in each FASTQ file were retained. The discarded sequences represent sequences with more errors, e.g., miscalls. The raw number of counts for each sample, presented in table format, can be found in [S1 Dataset](#).

Inherent bias in construction and finite reads meant that not every allele appeared in the naïve (unselected) cases. Thus enrichment in this study was defined as the relative increase in a particular allele after selection with respect to the population before antibiotic was applied (time point zero [TP0]) according to the following formula for enrichment of a given allele A.

$$\text{enrichment}_A = \left( \frac{\text{counts}_A}{\sum_n \text{counts}_n} \right)_{\text{selection}} - \left( \frac{\text{counts}_A}{\sum_n \text{counts}_n} \right)_{\text{TP0}}$$

The “top” alleles described are the alleles in each selection case with the highest enrichment over the naïve case.

## Bioinformatic Analysis

Most analyses were performed using the QIIME (Quantitative Insights into Microbial Ecology) pipeline, version 1.7.0 [32]. The open-source QIIME pipeline was built using the PyCogent libraries [33] and the Python programming language. Once installed, QIIME analyses are performed through a simple command-line interface, where the input and output file paths are specified, as well as any method parameters. QIIME was used for all of the following analyses: normalization, formation of a distance matrix, principal coordinates analysis (PCoA), Procrustes analysis, supervised learning, part of the network analysis, COG relative abundance plots, and ANOSIM. The QIIME scripts used for the above and below list of analyses were: `single_rarefaction.py`, `normalize_table.py`, `beta_diversity.py`, `principal_coordinates.py`, `transform_coordinate_matrices.py`, `supervised_learning.py`, `make_otu_network.py`, `summarize_taxa_through_plots.py`, and `compare_categories.py`. All of these QIIME scripts use as input the table of gene counts in each sample, and corresponding metadata, found in [S1 Dataset](#), or products from previously used scripts (e.g. `beta_diversity.py` should be used before `principal_coordinates.py`).

First, the tables in [S1 Dataset](#) were normalized. Normalization is necessary to correct for uneven library sizes, as well as other artifacts of the sequencing process [34]. The tables were subsampled (rarefied) to a depth of 2000 sequences per sample. Another normalization method implemented in R and QIIME, metagenomeSeq’s cumulative sum scaling (CSS), was performed in order to ensure robustness of results [35]. Next, a distance matrix was formed using Bray-Curtis dissimilarity [36, 37], since antibiotics selecting for the same genes should be deemed more similar, and because Bray-Curtis is less sensitive to data sparsity and compositionality [38–40]. Then, PCoA was performed on the distance matrices. We also assessed the results using Euclidean and binary Jaccard metrics with similar results.

Procrustes analysis, which enables comparison of the relative distances between points in two multivariate datasets, [41] was also performed on the gene and COG distance matrices. The measure of fit ( $M^2$ ) was calculated as the sum of the squared distances between corresponding sample points after the data is translated, rotated, and scaled to minimize the distance between the two datasets. The p-value was calculated by 1000 Monte-Carlo permutations in which the sample labels were randomly permuted; the number of iterations in which the  $M^2$  value was lower than the actual was divided by 1000 to yield the p-value.

Supervised learning was performed in QIIME on the two tables in [S1 Dataset](#) using the random forest machine learning method [42], with 5,000 sequences per sample, 500 trees, and leave-one-out cross-validation to estimate the generalization error and feature importance [43, 44]. Plots of alleles based on genomic location were generated using Circos software [45]. Genes were annotated with their corresponding Clusters of Orthologous Groups (COGs) [46]. Relativized counts were plotted using the `summarize_taxa` scripts in QIIME [32]. Networks

were constructed using Cytoscape [47]. ANOSIM was also carried out in QIIME [48] using the 'vegan' package in R [49].

## Results and Discussion

### Selection of antibiotic-resistant alleles from a genome-scale library

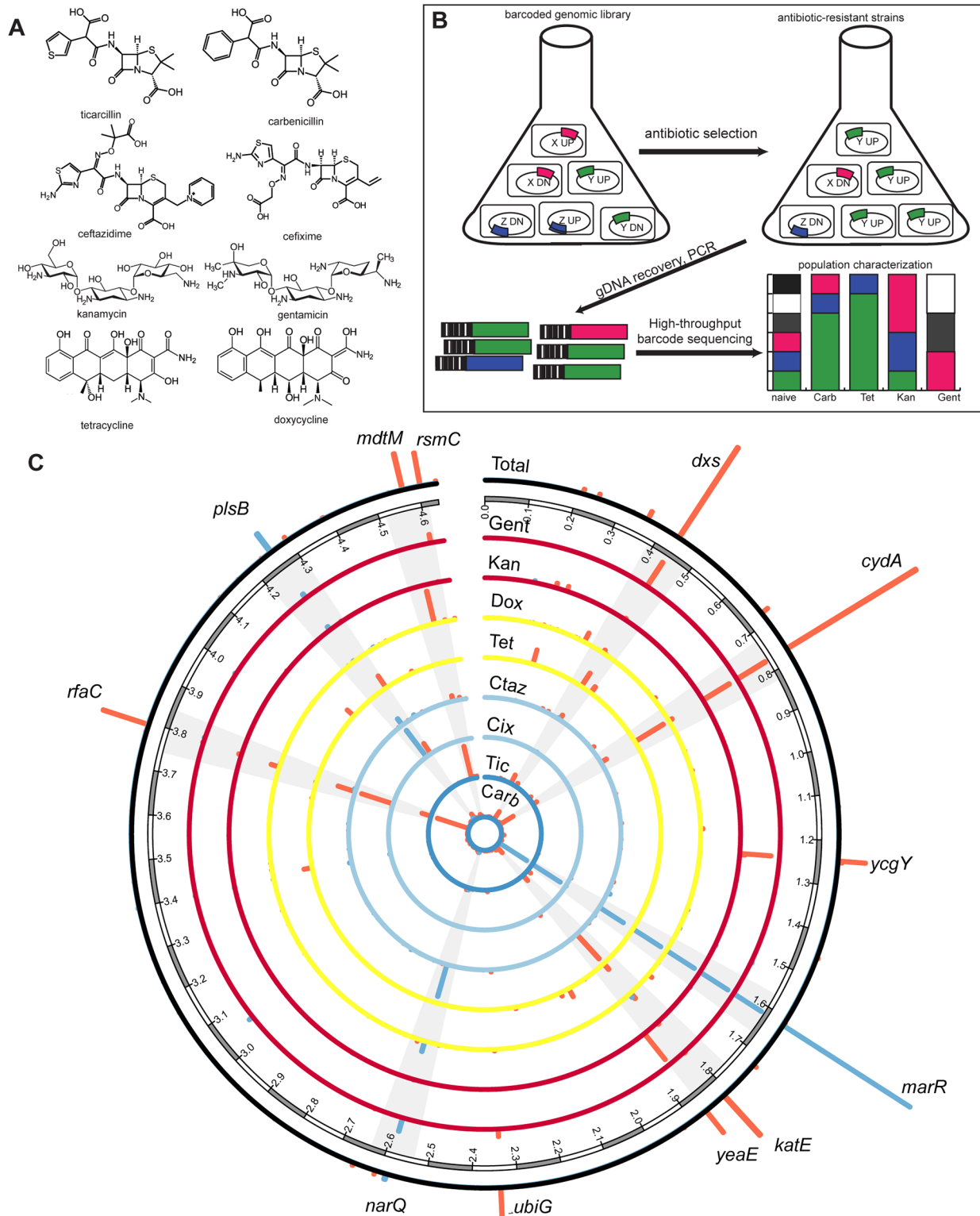
We subjected our genome-scale, barcoded library to selection on eight different antibiotics with three different mechanisms of action (S1 Table). Pairs of antibiotics were selected for chemical similarity (e.g., ticarcillin differs from carbenicillin only by the substitution of a five-membered thiophenyl moiety for a benzyl moiety) (Fig 1a). Briefly, *E. coli* MG1655 cells were previously subjected to multiplex recombineering using synthetic DNA cassettes containing either an "up" (strong promoter and ribosome binding sequence [RBS]) or "down" (no promoter or RBS) sequence along with homology regions corresponding to 4,077 genes in the *E. coli* genome. The synthetic cassettes also contained unique barcodes for rapid quantification of each of the approximately 8,000 TRMR mutants by microarray or pyrosequencing technologies (S1 Fig).

To design our growth selections, we first measured the minimum inhibitory concentration (MIC) for each antibiotic of interest in a strain equivalent to the parent strain of the TRMR library. MG1655-BC, a version of MG1655 with a barcode inserted at a silent site (the *attn7* site), was grown in liquid culture in triplicate at varying amounts of antibiotic to determine the concentration at which growth of the wild-type strain was inhibited (S1 Table). We were initially surprised by the MIC of MG1655-BC cells with respect to the  $\beta$ -lactam antibiotics (e.g., 44  $\mu\text{g}/\text{mL}$  for carbenicillin, given a typical working concentration of 50 to 100  $\mu\text{g}/\text{mL}$  for *E. coli*). However, further investigation of the literature revealed that the MIC of *E. coli* for carbenicillin is highly strain-dependent. According to the Antimicrobial Index [50], MICs of carbenicillin for strains of plasmid-free *E. coli* can measure from 1.56 to 32  $\mu\text{g}/\text{mL}$ . In addition, a recent report demonstrated that  $\beta$ -lactam resistance in MG1655 is particularly susceptible to sugar levels and inoculum [51], which was consistent throughout all experiments in this work. The MICs of all other antibiotics were within ranges that have been reported in literature for MG1655 [50].

Once the MIC was determined, the TRMR library was inoculated in triplicate in two identical sets of flasks containing MOPS rich defined media and one of eight antibiotics of interest at 80% of the MIC (48 flasks total). We performed selections at these concentrations in an attempt to normalize the selective pressure across all antibiotics. These flasks were grown until the late exponential phase with samples extracted at 24 hours and upon reaching late exponential phase. Genomic DNA was extracted and used as a template for PCR amplification and preparation for Illumina HiSeq sequencing of the barcode region (Fig 1b). More than 22 million barcode reads were counted and assigned to individual clones and degree of enrichment (see Methods) were calculated for all 8,077 TRMR mutants in each of the selections performed (Fig 1c). This analysis identified alleles enriched in TRMR libraries after selection that are consistent with previous studies on antibiotic resistance, plus uncharacterized genes potentially involved in resistance that could be important for further study (S2 Table). In addition, we use enrichment measurements to report alleles that may confer hypersensitivity (S3 Table, S2 Fig).

### Alleles contributing to antibiotic resistance

Our data suggest that multi-drug resistance alleles are consistently enriched regardless of the antibiotic selection performed (Fig 1c), and comprise a large fraction (10–90%) of each of the selected populations (S3 Fig). Specifically, we found five alleles that were enriched in all eight selections and six that were enriched in all but one case. These 11 alleles comprised over 30% of



**Fig 1. Selection of a genome-scale library on several antibiotics yields multi-drug resistant genes.** (a) Chemical structures of the eight antibiotics used in this study. (b) The TRMR library containing strains simulating “up” or “down” expression phenotypes in *E. coli* is grown in selective conditions. The genomic DNA of the survivors is harvested and amplified by PCR and the amplicon is sent to high-throughput sequencing, after which it is analyzed. (c) Enrichment scores for TRMR “up” (blue) or “down” (red) alleles for particular antibiotics are plotted relative to their location in the *E. coli* genome (in Mb). Alleles enriched in many or all selections are highlighted. The outside ring represents a linear combination of all eight antibiotic trials.

doi:10.1371/journal.pone.0146916.g001

the selected population in six cases, but comprised only 0.6% of the population before selection. These results suggest that laboratory selections enrich for MDR alleles (generalists), and not only for distinct sets of individual antibiotic resistance alleles (specialists). It is important to note that previous selections of the TRMR library on the same media without antibiotics [15] did not result in significant enrichment of any of the below noted alleles (i.e., all rank below the 100 most highly enriched in MOPS media alone).

One of the most prevalent alleles, occurring in the ten most highly enriched alleles in all cases (Fig 1c) is *marR*\_up. In this construct, the *marRAB* (where *mar* stands for “multiple antibiotic resistance”), which is normally negatively autoregulated by *marR* [52], is under control of the TRMR strong and constitutive promoter ( $p_{L_{tetO}}$ ). *MarA* is known to regulate several genes involved in resistance to antibiotics and multidrug efflux [53]. The *rfaC*\_down strain occurs in the ten most highly enriched alleles in seven of the eight cases. In this mutant (and all other TRMR “down” mutants), the native RBS has been removed to minimize translation. *RfaC* catalyzes a key step in lipopolysaccharide synthesis [54]. *RfaC* mutants in several pathogenic bacteria including *E. coli* show increased resistance to various antibiotics [55]. It is not clear why the “down” mutation was selected (as opposed to the “up” mutation). However, because the blasticidin resistance cassette contains a strong EM7 promoter 5’ of the gene of interest (S1 Fig), it is possible that some read-through may occur, leading to constitutive downstream expression.

Other alleles consistently enriched by selection with several antibiotics and previously associated with antibiotic resistance included genes related to (1) managing oxidative stress: *katE* [56] and *sodC* [57], (2) oxidative phosphorylation: *cydA* [10], (3) transport and efflux: *mdtM* [58], and (4) other metabolic processes: *dxs* [59], and *plsB* [60]. We then confirmed that apparent increased antibiotic resistance led to increased growth on many antibiotics. S4 Fig shows the 24 hour OD<sub>600</sub> on various concentrations of all antibiotics for several isolated individual strains with their enrichment scores for comparison (S4 Fig). In addition to genes enriched in multiple selections, we identified a range of genes enriched in individual selections and several genes of unknown or uncharacterized function (S2 Table). The Additional Data file lists allele counts for all selections in this work.

Among the genes enriched in individual selections was *rsmC*, a 16S ribosomal subunit nucleotide methylase. The *rsmC*\_down allele was highly enriched in the gentamicin selection. Interestingly, a recent study implicates 16S ribosomal RNA methylases in aminoglycoside resistance in *Enterobacteriaceae* [61]. A highly enriched allele for cefixime resistance was *mreC*\_up. *MreC* is a rod-shape determining protein involved in peptidoglycan synthesis that has been associated with  $\beta$ -lactam resistance in *Helicobacter pylori* [62]. *SecD*, another allele isolated in the cefixime selection, has also been linked to  $\beta$ -lactam resistance in *E. coli* [55].

Unexpectedly, several enriched alleles for cefixime, ticarcillin, and gentamicin selection(s) corresponded to hydrogen production and formate processing including *fdnG*, *hyfJ*, and *narQ*. It is possible that the actions of these proteins affect the proton motive force, either facilitating increased drug efflux by increased PMF or decreasing drug uptake by reducing PMF (as is well known to affect the toxicity of charged compounds such as aminoglycoside antibiotics [63]). Several alleles were isolated that correspond to genes with unknown functions. They include: *ycjO* (putative ABC transporter), *yiiR*, *ybeT* (conserved outer membrane protein), *yafL* (inner membrane protein), *ycgY*, *yeaE* (methylglyoxal reductase), *yebY*, *yigB*, *yiiR*, and *yncH*. The contribution of these genes to antibiotic resistance warrants further investigation.

Finally, the targeting of antibiotic sensitivity genes provides a possible mechanism to treat resistant infections. To determine genes that might convey sensitivity to the antibiotics of interest, we also recorded the alleles with the lowest degree of enrichment (i.e., largest decrease in frequency throughout a selection) (S3 Table). Our analysis suggested considerable overlap in

susceptibility genes across the antibiotics investigated (S2 Fig). Many of the proteins encoded by these alleles are targeted to the inner membrane. Previous experiments also showed that these specific alleles grew well on non-selective rich MOPS media [15]. Thus, it is possible that changes in expression of these inner membrane proteins alter the overall inner membrane fluidity or porosity, allowing antibiotics to traverse membrane more easily. While this possibility should not be discounted, it should also be noted that all of the above susceptible alleles were present in large quantities at time point zero. Given the strength of each selection, it is possible that these alleles were simply diluted down to the limit of detection. This is an issue of selection design; in particular our designs were targeted at enrichment for resistance phenotypes as opposed to identification of susceptibility phenotypes.

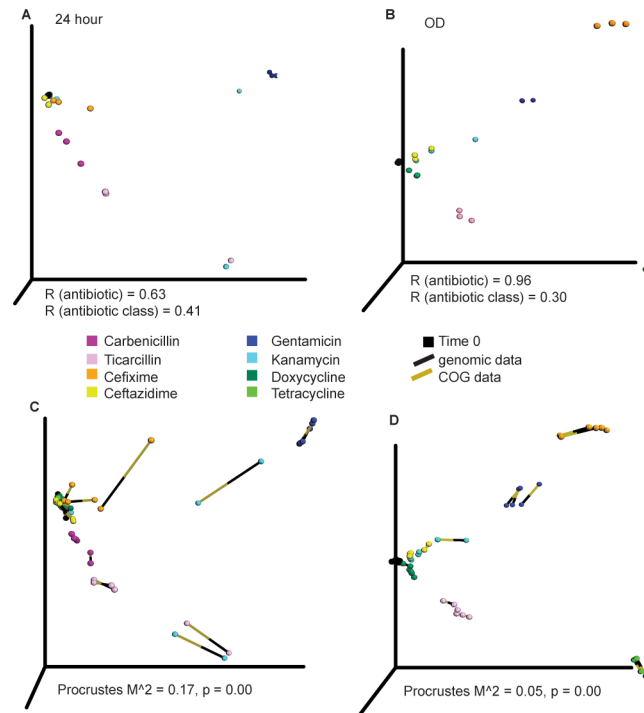
### Allelic responses to chemically similar antibiotics are weakly dissimilar

Our data suggested that sub-lethal antibiotic treatment strategies selected for multi-drug resistance alleles. To explore this suggestion in more depth, we performed principal coordinate analysis (PCoA) on all replicates from each selection. PCoA allows for visualization of multi-dimensional variables in 3D space by condensing distance metrics into the most important coordinates while minimizing the loss of information. We specifically hypothesized that antibiotics with similar chemical structure and belonging to the same class (e.g., gentamicin and kanamycin) would present a similar allelic response and therefore cluster together in PCoA space, and that antibiotics having similar mechanisms of action (e.g., the aminoglycosides and the tetracyclines, which both act by binding 30S ribosomal subunits) would as well. Although some patterns appear at 24 hours and after reaching the late exponential phase (Fig 2a and 2b), such as the location of gentamicin and kanamycin in the upper half of PCoA space, other patterns are unexplained. For example, doxycycline, carbenicillin, and ceftazidime cluster near time point zero. This finding is supported by a weaker ANOSIM R value for antibiotic class or mechanism of action (Fig 2a and 2b). ANOSIM R-values near zero indicate random grouping. Network analysis, in which samples sharing similar genes are drawn together, confirms that the subtle antibiotic PCoA clustering patterns, as there are no large differences between antibiotic types (S5 Fig). However, differences between antibiotics are discernable by ANOSIM [64], which is an extremely sensitive test (S4 Table). These results are robust to normalization technique, replicate, and distance metric (S6 and S7 Figs).

### Clusters of Orthologous Groups analysis elucidates functional hierarchy

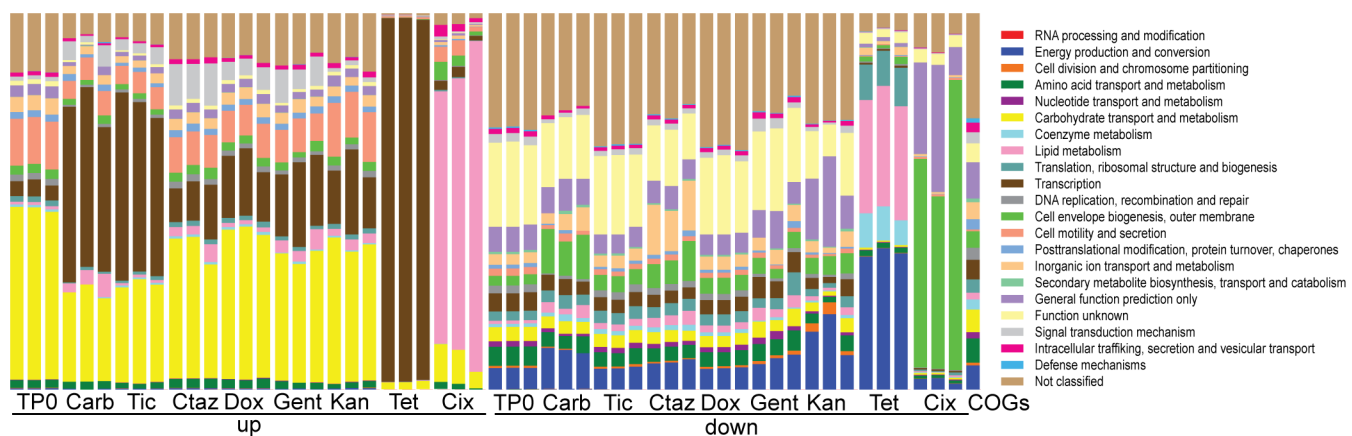
Although antibiotics of similar classes or targeting the same complex did not exhibit significant clustering in PCoA space at the specific allele level, we speculated that clearer patterns might be revealed when the PCoA analysis was performed at the level of encoded functions. To gain an understanding of mechanisms of action on a pathway level [65], a matrix of clusters of orthologous groups (COG) [46] was formed by summing the counts of genes belonging to the same COG in the same sample. We then performed Procrustes analysis to analyze the similarity of the gene and COG distributions in PCoA space (Fig 2c and 2d). Procrustes analysis stretches, rotates, and scales two datasets to determine if similar conclusions could be drawn [66]. The p-values are less than 0.001, suggesting that the functional profiles could be predicted from the TRMR alleles enriched by selection because both matrices display similar PCoA clustering patterns. This match between COG and gene distributions implies that selection acts in broadly similar ways at multiple levels of the functional hierarchy. Fig 3 shows the similar COG distribution of the antibiotic samples over time. However, the “up” and “down” alleles perform very different functions, and are different from the wild-type distribution (far right bar). The TRMR “up” alleles were enriched in COG categories of carbohydrate metabolism and transport





**Fig 2. Separation of antibiotic classes in PCoA space is weak across multiple levels of functional hierarchy.** (a,b) PCoA analysis, using Bray-Curtis distance, of the antibiotics at (a) 24 hours (b) upon reaching the late exponential phase (OD). ANOSIM R-values are plotted for separation by antibiotic or by mechanism of action. (c,d) Procrustes analysis indicates significant alignment between the COG (gold end of the line) and gene (black end of the line) PCoA profiles in the 24 h and OD selections. The longer the line connecting the COG and gene points, the less aligned the two points are in PCoA space, increasing the stress value ( $M^2$ ).

doi:10.1371/journal.pone.0146916.g002



**Fig 3. Clusters of Orthologous Groups (COGs) relative abundance of selected populations.** Label format is Antibiotic\_replicate\_up/down, e.g., CarbOD\_1\_up means COG counts for carbenicillin, replicate 1 of 3, and TRMR “up” alleles. Antibiotic abbreviations: Carbenicillin (Carb), Ticarcillin (Tic), Ceftazidime (Ctaz), Cefixime (Cix), Gentamicin (Gent), Kanamycin (Kan), Doxycycline (Dox), and Tetracycline (Tet). Far right bar: COG distribution as represented in the wild-type *E. coli* genome.

doi:10.1371/journal.pone.0146916.g003

(yellow), and transcription (brown). The latter was expected for the tetracyclines and aminoglycosides as ribosomal inhibitors, but not for  $\beta$ -lactams. The functions of the TRMR “down” alleles were much more diverse.

## Supervised learning distinguishes resistance “fingerprints”

Given that MDR alleles were a significant fraction of every selection (S3 Fig), we wanted to understand whether the final antibiotic populations could be distinguished. To do so, we used supervised learning to identify combinations of genes that may be unique to individual antibiotics, and thus represent a genomic “resistance fingerprint” for each antibiotic. We used the random forest classifier [42] to generate confusion matrices from 48 samples (24 hour and late-exponential phase selections, in triplicate, on each of eight antibiotics), which indicate true vs. predicted classifications when a portion of the dataset is withheld from model training (Fig 4). At the level of individual alleles, it was difficult to distinguish between some antibiotics (as shown by shading off of the diagonal), especially between antibiotics of the same class or mechanism of action (Fig 4a). The random forest classifier returns a ratio of baseline error to observed error of 2.2, indicating that the classifications are estimated to be 2.2 times more accurate than random guessing, a statistically significant but weak effect.

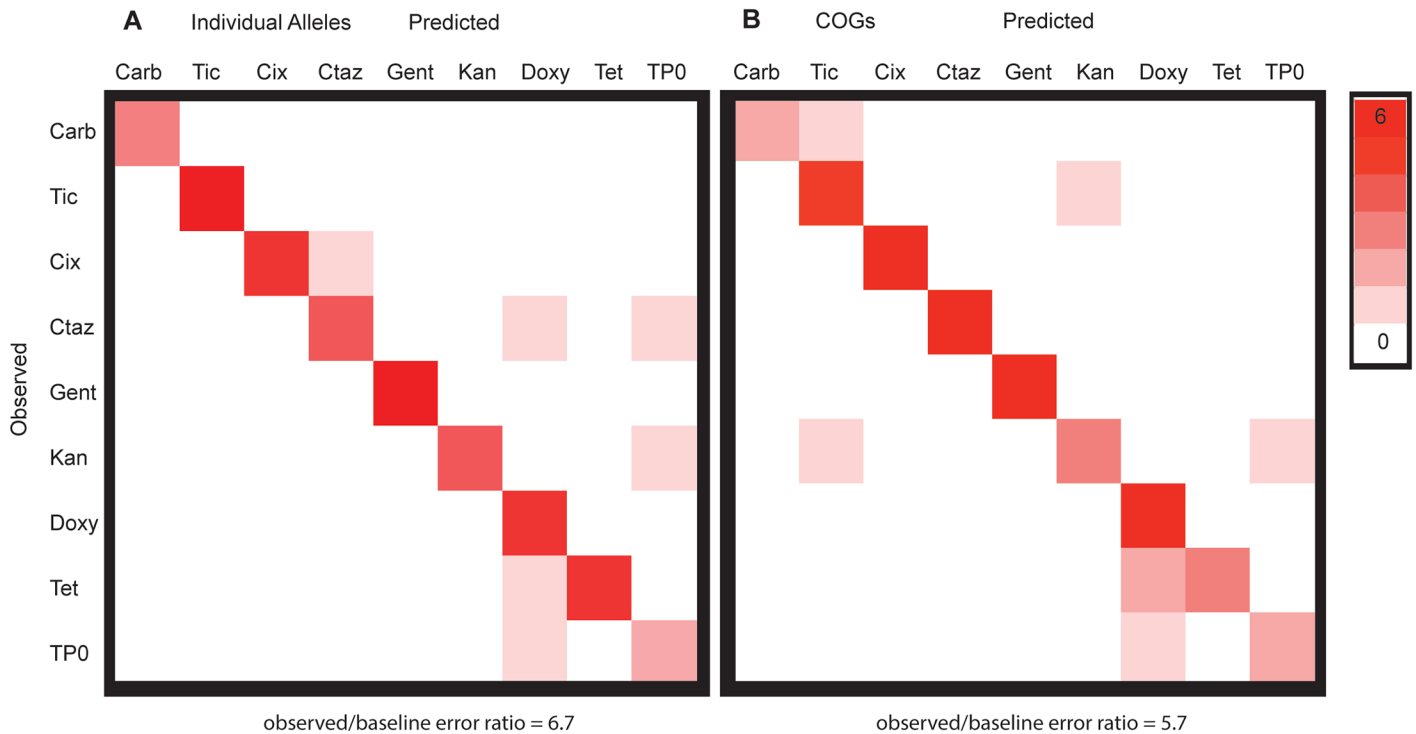
However, when alleles are grouped by COG category, supervised learning improves substantially. There is excellent classification of antibiotics with a baseline error ratio of 5.7 (Fig 4b). This indicates that each antibiotic has a unique signature at the COG level. Classification between the antibiotics may further improve by adding more antibiotics within each class. This hypothesis is supported by perfect distinction (baseline error ratio 24.0) between gentamicin and kanamycin, antibiotics of the same class with similar chemical structure, in a separate detailed time course selection (S8 Fig). Furthermore, the subtle genetic differences arise at the first antibiotic application, independent of selection length (S8 Fig).

The genes, COGs, and the enrichment patterns the random forest classifier uses most to distinguish between the antibiotics are found in S9 and S10 Figs. Interestingly, most of the genes that are key in building the antibiotic classifier, which examines the prediction strengths of individual genes, are also identified as the high/low enrichment alleles analyzed in the above genomic plots (Fig 1a, S2 Fig). Also, the distinction between antibiotics and their classes diminishes when using the binary Jaccard distance metric, which operates on a gene presence/absence basis (S7 Fig). This strengthens the conclusion that while alleles conferring multi-drug resistance are found in many cases, variation in the degree of enrichment of these MDR alleles for a particular antibiotic is a predictor of the genetic fingerprint of a particular antibiotic.

## Conclusions

We have presented a model pipeline for the analysis of gene products leading to antimicrobial resistance in *E. coli*. We discovered that many alleles isolated from treatment with low levels of single antibiotics conferred resistance to many antibiotics. This lends support to the hypothesis that low-dose antibiotics as used in livestock growth promotion and found in wastewater likely promote resistance to a wide range of antimicrobial compounds including last-resort therapeutics [17]. The rise of antimicrobial resistance is also important in microbial ecology, including soil [67] and human gut [68] bacteria.

Chemical tolerance in microbes is often a complex phenotype conferred by a range of genetic factors that are often not intuitive or obvious. A seminal work in chemical genomics in *E. coli* was recently published in which a library of over 4,000 strains including the Keio deletion library was screened under many different chemical and physical conditions [3]. In that work, individual strains were plated robotically in 1,536-well format, and colony size was



**Fig 4. Supervised learning is able to distinguish between the antibiotics at both the allele and COG level.** Confusion matrices for random forest classifiers. Off diagonal classification represents classifier error. Antibiotic abbreviations as in Fig 3.

doi:10.1371/journal.pone.0146916.g004

investigated to determine fitness. A similar work examined the effect of a library of 4,000 *E. coli* genes overexpressed on plasmids challenged by a variety of chemicals [25]. While the library was assayed in multiplex in microtiter plates, characterization of alleles (by the nature of Sanger sequencing) was limited to less than 10 colonies per condition. A similar study of all Keio collection knockouts on 14 different antibiotics also required robotics and was focused on antibiotic sensitivity [26]. This study discovered multiple-resistance knockouts in the Keio collection. While there was little overlap with our data, of the 14 antibiotics used, only two were used in this study, highlighting the complexity of multi-drug resistance. A recent study focused on aminoglycoside antibiotics used adaptive evolution over hundreds of generations to examine beneficial mutations and characterized by whole genome sequencing of 240 parallel-evolved lines [6]. Several of the above-mentioned studies concluded that mutations that affected efflux pumps such as AcrAB contributed to multiple-drug resistance. Our observation of the *marR* allele observed agrees with this result, but as the scope of our search was much broader we were also able to determine multi-drug resistant alleles with mechanisms which do not necessarily have to do with efflux pump regulation as well as alleles with unknown function without whole-genome sequencing. While we did not see significant enrichment of *acrAB\_up*, it is possible that overexpression of these membrane proteins is more detrimental to the cell than the efflux of the antibiotic, resulting in an overall slower growth rate.

The original application of the TRMR library used DNA microarrays and exploratory, not multivariate, analyses to characterize the genome-level responses to various conditions. However, this application required custom-made arrays corresponding to the barcodes. In addition, as demonstrated by the application of Bar-seq to a yeast deletion library [14], sequencing has many advantages over microarrays for rapid analysis of barcoded libraries, including but not

limited to cost, the ability to pursue many biological replicates under various conditions in one sequencing lane, reduced crosstalk, and increased resolution on the low and high ends of detection [69]. Previously, DNA sequencing data from barcoded libraries was analyzed using packages with specialized analyses, and for smaller, number dense datasets [70]. In contrast, sparse datasets (containing many zeroes) like the one presented in this work make metagenomic techniques like the analyses in QIIME more appropriate [35]. QIIME also contains many analysis types in one package, streamlining analyses, and can easily analyze dataset sizes from small to massive [71]. As input, QIIME only requires sample metadata (e.g., antibiotic type) and a large count matrix derived from high-throughput sequencing, in which samples are the columns, and the counts of each feature (gene, microbe, etc.) are the rows.

Overall, our approach allows such analyses in multiplex at the level of growth selections (over roughly 24 to 48 hours) and now in the sequencing steps as well, allowing considerably faster, deeper, and larger laboratory population genomic dynamics studies in bacteria. Barcoding maximizes the usefulness of short reads and allows for the use of HiSeq technology to generate millions of times more data points than Sanger sequencing would allow. In addition, the barcoded and pre-defined nature of the library circumvents the need for long adaptation cycles (10–100 times fewer generations required) and whole genome sequencing. Thus, the combination of a method to map the specific effect of genes to selectable traits (TRMR), high-throughput sequencing, and streamlined bioinformatics analysis software (QIIME) provides a powerful toolbox for exploring the genetic basis of a broad variety of complex phenotypes [32]. Finally, the same methodologies of selection, high-throughput sequencing, and bioinformatic analysis are broadly applicable to experiments on chemical tolerance for any inhibitory chemical, from antibiotics to toxic metabolites to next-generation biofuels.

## Supporting Information

### S1 Dataset. Raw counts and enrichment scores.

(XLSX)

### S1 Fig. Schematic of the inserted cassette in TRMR library mutants.

(JPG)

**S2 Fig. Multi-drug sensitivity genes selected for across all eight antibiotics.** TRMR “up” (blue) or “down” (red) alleles conveying the lowest enrichment for particular antibiotics are plotted relative to their location in the *E. coli* genome (in Mb). Alleles diminished in many or all selections are highlighted. The outer ring represents a linear combination of all eight antibiotic trials.

(JPG)

### S3 Fig. Percent of selected populations comprising multi-drug resistant genes.

(JPG)

### S4 Fig. $OD_{600}$ of liquid cultures of individual TRMR clones at various antibiotic levels.

Heat map representing average optical density of triplicate cultures in MOPS-glucose media. Eight antibiotic concentrations were used at two-fold serial dilutions. Maximum concentrations in  $\mu\text{g/mL}$ : gentamicin: 1.52, kanamycin: 30, tetracycline: 18, doxycycline: 18, carbenicillin: 177.6, ticarcillin: 124.8, cefixime: 6.5, ceftazidime: 13. Overlaid in white: Enrichment scores for each clone on each antibiotic.

(TIF)

**S5 Fig. Network analysis of antibiotic selections.** (A) Nodes are the antibiotic type, while the black dots are the genes. If a gene is shared between the antibiotics, it pulls those nodes closer

at an amount weighted by the gene's abundance. If a gene is not shared between the antibiotics, it pulls the antibiotic sample node it is attached to towards the outside of the diagram, separating the nodes. The close clustering of the antibiotic nodes indicates many shared genes. (B) The separate clustering of the TRMR 'up' vs. the TRMR 'down' antibiotic selections indicates that very different up/down genes are selected for.

(JPG)

**S6 Fig. Clustering by antibiotic class is consistent regardless of normalization technique.**

24 hour time point (left) and late exponential phase (OD) selections (right). The rows are the normalization methods used, which are rarefying or cumulative-sum scaling (CSS) (Paulson et al., Nature Methods, 2013).

(JPG)

**S7 Fig. ANOSIM R-values are consistent regardless of distance metrics.** 24 hour (left) and late exponential phase (OD) selections (right). Each row represents clustering with a different distance metric. The much smaller ANOSIM R-value for the binary Jaccard selections supports the hypothesis of [S3 Fig](#): that differences in allelic population abundances, rather than the alleles themselves, are the main variable driving the antibiotic separation.

(JPG)

**S8 Fig. A detailed time course selection, with many more samples, on gentamicin and kanamycin results in near perfect supervised learning classification.**

(A) Schematic of the types of time-course experimental setups. All of these selections were done in triplicate to control for experimental variations. (B) The majority of change to the allele population occurs in the first selection, regardless of selection type. This is shown by the large Bray-Curtis distance between Time\_0 (TP0) and the selection types. Also, the second and third constant transfers (CT2, CT3) or serial dilutions (SD2, SD3) do not have much higher bars than CT1 or ST1. (C) Supervised learning confusion matrix for the detailed Gentamicin and Kanamycin time course study shows no error (off diagonal classification) between the two antibiotics.

(JPG)

**S9 Fig. Heatmap of the log<sub>10</sub> counts of the top 25 genes that distinguish antibiotic categories in the supervised learning classifier.**

Label format is Antibiotic/ Method\_replicate. For example, Carb24\_1 means Carbenicillin was used, it is the 24- hour selection, and it is replicate 1 of 3.

(TIF)

**S10 Fig. Heatmap of the log<sub>10</sub> counts of the COG categories used in the supervised learning classifier.**

Labeling as in [S9 Fig](#). COG category symbol and meaning: C—Energy production and conversion, D—Cell cycle control and mitosis, E—Amino Acid metabolism and transport, F—Nucleotide metabolism and transport, G—Carbohydrate metabolism and transport, H—Coenzyme metabolism, I—Lipid metabolism, J—Translation, K—Transcription, L—Replication and repair, M—Cell wall/membrane/envelope biogenesis, N—Cell motility, O—Post-translational modification, protein turnover, chaperone functions, P—Inorganic ion transport and metabolism, Q—Secondary structure, T—Signal transduction, U—Intracellular trafficking and secretion, Y—Nuclear structure, Z—Cytoskeleton, R—General function prediction only, S—Function unknown.

(TIF)

**S1 Table. Antibiotics and concentrations used in this study.**

(XLSX)

**S2 Table. Top 10 highly enriched alleles in OD selections for each antibiotic.**  
(XLSX)

**S3 Table. Top 10 diminished alleles in OD selections for each antibiotic.**  
(XLSX)

**S4 Table. Nonparametric ANOSIM values for important categories in this study.** The R statistic represents the how different the tested categories are, with a value near zero indicating no significant difference between the groups, and a value near 1 indicating difference.  
(XLSX)

**S5 Table. A1 top genes used for classification between the eight antibiotics.**  
(XLS)

## Acknowledgments

The authors thank Anis Karimpour-Fard for providing the COG gene assignments, and Tirzah Glebes and Lauren Woodruff for helpful discussions. SJW was funded by the NIH/CU Molecular Biophysics Training Scholarship (T32 GM-065103).

## Author Contributions

Conceived and designed the experiments: SJW RTG. Performed the experiments: SJW TJM. Analyzed the data: SJW TJM. Contributed reagents/materials/analysis tools: PM. Wrote the paper: TJM SJW RK RTG. Oversaw experimental design: TJM RTG. Oversaw data analysis: RK. Supervised all aspects of the project: RTG.

## References

1. Hillenmeyer ME, Ericson E, Davis RW, Nislow C, Koller D, Giaever G. Systematic analysis of genome-wide fitness data in yeast reveals novel gene function and drug action. *Genome Biol.* 2010; 11(3):R30. doi: [10.1186/gb-2010-11-3-r30](https://doi.org/10.1186/gb-2010-11-3-r30) PMID: [20226027](https://pubmed.ncbi.nlm.nih.gov/20226027/)
2. Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, et al. The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science.* 2008; 320(5874):362–5. doi: [10.1126/science.1150021](https://doi.org/10.1126/science.1150021) PMID: [18420932](https://pubmed.ncbi.nlm.nih.gov/18420932/)
3. Nichols RJ, Sen S, Choo YJ, Beltrao P, Zietek M, Chaba R, et al. Phenotypic landscape of a bacterial cell. *Cell.* 2011; 144(1):143–56. doi: [10.1016/j.cell.2010.11.052](https://doi.org/10.1016/j.cell.2010.11.052) PMID: [21185072](https://pubmed.ncbi.nlm.nih.gov/21185072/)
4. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, et al. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science.* 1999; 285(5429):901–6. PMID: [10436161](https://pubmed.ncbi.nlm.nih.gov/10436161/)
5. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol.* 2006; 2:2006.0008. PMID: [16738554](https://pubmed.ncbi.nlm.nih.gov/16738554/)
6. Lazar V, Pal Singh G, Spohn R, Nagy I, Horvath B, Hrtyan M, et al. Bacterial evolution of antibiotic hypersensitivity. *Mol Syst Biol.* 2013; 9:700. doi: [10.1038/msb.2013.57](https://doi.org/10.1038/msb.2013.57) PMID: [24169403](https://pubmed.ncbi.nlm.nih.gov/24169403/)
7. Toprak E, Veres A, Yildiz S, Pedraza JM, Chait R, Paulsson J, et al. Building a morbidostat: an automated continuous-culture device for studying bacterial drug resistance under dynamically sustained drug inhibition. *Nat Protoc.* 2013; 8(3):555–67. doi: [10.1038/nprot.nprot.2013.021](https://doi.org/10.1038/nprot.nprot.2013.021) PMID: [23429717](https://pubmed.ncbi.nlm.nih.gov/23429717/)
8. Kosuri S, Church GM. Large-scale de novo DNA synthesis: technologies and applications. *Nat Methods.* 2014; 11(5):499–507. doi: [10.1038/nmeth.2918](https://doi.org/10.1038/nmeth.2918) PMID: [24781323](https://pubmed.ncbi.nlm.nih.gov/24781323/)
9. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet.* 2014; 15(2):121–32. doi: [10.1038/nrg3642](https://doi.org/10.1038/nrg3642) PMID: [24434847](https://pubmed.ncbi.nlm.nih.gov/24434847/)
10. Gill RT, Wildt S, Yang YT, Ziesman S, Stephanopoulos G. Genome-wide screening for trait conferring genes using DNA microarrays. *Proc Natl Acad Sci U S A.* 2002; 99(10):7033–8. PMID: [11997466](https://pubmed.ncbi.nlm.nih.gov/11997466/)
11. Lynch MD, Warnecke T, Gill RT. SCALEs: multiscale analysis of library enrichment. *Nat Methods.* 2007; 4(1):87–93. PMID: [17099705](https://pubmed.ncbi.nlm.nih.gov/17099705/)

12. Jin YS, Stephanopoulos G. Multi-dimensional gene target search for improving lycopene biosynthesis in *Escherichia coli*. *Metab Eng*. 2007; 9(4):337–47. PMID: [17509919](#)
13. Mazurkiewicz P, Tang CM, Boone C, Holden DW. Signature-tagged mutagenesis: barcoding mutants for genome-wide screens. *Nat Rev Genet*. 2006; 7(12):929–39. PMID: [17139324](#)
14. Smith AM, Heisler LE, Mellor J, Kaper F, Thompson MJ, Chee M, et al. Quantitative phenotyping via deep barcode sequencing. *Genome Res*. 2009; 19(10):1836–42. doi: [10.1101/gr.093955.109](#) PMID: [19622793](#)
15. Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LB, Gill RT. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat Biotechnol*. 2010; 28(8):856–62. doi: [10.1038/nbt.1653](#) PMID: [20639866](#)
16. Sandoval NR, Kim JY, Glebes TY, Reeder PJ, Aucoin HR, Warner JR, et al. Strategy for directing combinatorial genome engineering in *Escherichia coli*. *Proc Natl Acad Sci U S A*. 2012; 109(26):10540–5. doi: [10.1073/pnas.1206299109](#) PMID: [22689973](#)
17. Andersson DI, Hughes D. Evolution of antibiotic resistance at non-lethal drug concentrations. *Drug Resist Updat*. 2012; 15(3):162–72. doi: [10.1016/j.drug.2012.03.005](#) PMID: [22516308](#)
18. Andersson DI, Hughes D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat Rev Microbiol*. 2010; 8(4):260–71. doi: [10.1038/nrmicro2319](#) PMID: [20208551](#)
19. Diep BA, Gill SR, Chang RF, Phan TH, Chen JH, Davidson MG, et al. Complete genome sequence of USA300, an epidemic clone of community-acquired methicillin-resistant *Staphylococcus aureus*. *Lancet*. 2006; 367(9512):731–9. PMID: [16517273](#)
20. Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, et al. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet*. 2012; 44(1):106–10.
21. Xu C, Lin X, Ren H, Zhang Y, Wang S, Peng X. Analysis of outer membrane proteome of *Escherichia coli* related to resistance to ampicillin and tetracycline. *Proteomics*. 2006; 6(2):462–73. PMID: [16372265](#)
22. Piras C, Soggiu A, Bonizzi L, Gaviraghi A, Deriu F, De Martino L, et al. Comparative proteomics to evaluate multi drug resistance in *Escherichia coli*. *Mol Biosyst*. 2012; 8(4):1060–7. doi: [10.1039/c1mb05385j](#) PMID: [22120138](#)
23. Karatzas KA, Randall LP, Webber M, Piddock LJ, Humphrey TJ, Woodward MJ, et al. Phenotypic and proteomic characterization of multiply antibiotic-resistant variants of *Salmonella enterica* serovar Typhimurium selected following exposure to disinfectants. *Appl Environ Microbiol*. 2008; 74(5):1508–16. PMID: [18083849](#)
24. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, et al. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*. 2002; 418(6896):387–91. PMID: [12140549](#)
25. Soo VW, Hanson-Manful P, Patrick WM. Artificial gene amplification reveals an abundance of promiscuous resistance determinants in *Escherichia coli*. *Proc Natl Acad Sci U S A*. 2011; 108(4):1484–9. doi: [10.1073/pnas.1012108108](#) PMID: [21173244](#)
26. Liu A, Tran L, Becket E, Lee K, Chinn L, Park E, et al. Antibiotic sensitivity profiles determined with an *Escherichia coli* gene knockout collection: generating an antibiotic bar code. *Antimicrob Agents Chemother*. 2010; 54(4):1393–403. doi: [10.1128/AAC.00906-09](#) PMID: [20065048](#)
27. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*. 2006; 2:2006.0008. PMID: [16738554](#)
28. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A*. 2000; 97(12):6640–5. PMID: [10829079](#)
29. Neidhardt FC, Bloch PL, Smith DF. Culture medium for enterobacteria. *J Bacteriol*. 1974; 119(3):736–47. PMID: [4604283](#)
30. Andrews JM. Determination of minimum inhibitory concentrations. *J Antimicrob Chemother*. 2001; 48 Suppl 1:5–16. PMID: [11420333](#)
31. Hiatt JB, Patwardhan RP, Turner EH, Lee C, Shendure J. Parallel, tag-directed assembly of locally derived short sequence reads. *Nat Methods*. 2010; 7(2):119–22. doi: [10.1038/nmeth.1416](#) PMID: [20081835](#)
32. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nature methods*. 2010; 7(5):335–6. doi: [10.1038/nmeth.f.303](#) PMID: [20383131](#)
33. Knight R, Maxwell P, Birmingham A, Carnes J, Caporaso JG, Easton BC, et al. PyCogent: a toolkit for making sense from sequence. *Genome Biol*. 2007; 8(8):R171. PMID: [17708774](#)

34. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 2014; 12:87. doi: [10.1186/s12915-014-0087-z](https://doi.org/10.1186/s12915-014-0087-z) PMID: [25387460](https://pubmed.ncbi.nlm.nih.gov/25387460/)
35. Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. *Nat Methods.* 2013; 10(12):1200–2. doi: [10.1038/nmeth.2658](https://doi.org/10.1038/nmeth.2658) PMID: [24076764](https://pubmed.ncbi.nlm.nih.gov/24076764/)
36. Curtis JRB. An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs.* 1957; 27(4):325–49.
37. Bloom SA. Similarity indices in community studies: potential pitfalls. *Marine Ecology—Progress Series.* 1981; 5:125–8.
38. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, et al. Microbial co-occurrence relationships in the human microbiome. *PLoS computational biology.* 2012; 8(7):e1002606. doi: [10.1371/journal.pcbi.1002606](https://doi.org/10.1371/journal.pcbi.1002606) PMID: [22807668](https://pubmed.ncbi.nlm.nih.gov/22807668/)
39. Aitchison J. *The statistical analysis of compositional data.* London; New York: Chapman and Hall; 1986. xv, 416 p. p.
40. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. *PLoS computational biology.* 2012; 8(9):e1002687. doi: [10.1371/journal.pcbi.1002687](https://doi.org/10.1371/journal.pcbi.1002687) PMID: [23028285](https://pubmed.ncbi.nlm.nih.gov/23028285/)
41. Hurley JR, Cattell RB. The Procrustes program: Producing direct rotation to test a hypothesized factor structure. *Behavioral Science.* 1962; 7(2):258–62.
42. Breiman L. Random forests. *Machine learning.* 2001; 45(1):5–32.
43. Knights D, Costello EK, Knight R. Supervised classification of human microbiota. *FEMS Microbiol Rev.* 2011; 35(2):343–59. doi: [10.1111/j.1574-6976.2010.00251.x](https://doi.org/10.1111/j.1574-6976.2010.00251.x) PMID: [21039646](https://pubmed.ncbi.nlm.nih.gov/21039646/)
44. Knights D, Kuczynski J, Koren O, Ley RE, Field D, Knight R, et al. Supervised classification of microbiota mitigates mislabeling errors. *The ISME journal.* 2011; 5(4):570–3. doi: [10.1038/ismej.2010.148](https://doi.org/10.1038/ismej.2010.148) PMID: [20927137](https://pubmed.ncbi.nlm.nih.gov/20927137/)
45. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009; 19(9):1639–45. doi: [10.1101/gr.092759.109](https://doi.org/10.1101/gr.092759.109) PMID: [19541911](https://pubmed.ncbi.nlm.nih.gov/19541911/)
46. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics.* 2003; 4:41. PMID: [12969510](https://pubmed.ncbi.nlm.nih.gov/12969510/)
47. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003; 13(11):2498–504. PMID: [14597658](https://pubmed.ncbi.nlm.nih.gov/14597658/)
48. Clarke KR. Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology.* 1993; 118(1):117–43.
49. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci.* 2003; 14(6):927–30.
50. Thorsing M, Bentin T, Givskov M, Tolker-Nielsen T, Goltermann L. The bactericidal activity of beta-lactam antibiotics is increased by metabolizable sugar species. *Microbiology.* 2015; 161(10):1999–2007. doi: [10.1099/mic.0.000152](https://doi.org/10.1099/mic.0.000152) PMID: [26243263](https://pubmed.ncbi.nlm.nih.gov/26243263/)
51. Amirkia VD, Qiubao P. The Antimicrobial Index: a comprehensive literature-based antimicrobial database and reference work. *Bioinformatics.* 2011; 5(8):365–6. PMID: [21383924](https://pubmed.ncbi.nlm.nih.gov/21383924/)
52. Martin RG, Rosner JL. Transcriptional and translational regulation of the marRAB multiple antibiotic resistance operon in *Escherichia coli*. *Mol Microbiol.* 2004; 53(1):183–91. PMID: [15225313](https://pubmed.ncbi.nlm.nih.gov/15225313/)
53. Ruiz C, Levy SB. Many chromosomal genes modulate MarA-mediated multidrug resistance in *Escherichia coli*. *Antimicrob Agents Chemother.* 2010; 54(5):2125–34. doi: [10.1128/AAC.01420-09](https://doi.org/10.1128/AAC.01420-09) PMID: [20211899](https://pubmed.ncbi.nlm.nih.gov/20211899/)
54. Kadmas JL, Raetz CR. Enzymatic synthesis of lipopolysaccharide in *Escherichia coli*. Purification and properties of heptosyltransferase I. *J Biol Chem.* 1998; 273(5):2799–807. PMID: [9446588](https://pubmed.ncbi.nlm.nih.gov/9446588/)
55. Tamaki S, Sato T, Matsuhashi M. Role of lipopolysaccharides in antibiotic resistance and bacteriophage adsorption of *Escherichia coli* K-12. *J Bacteriol.* 1971; 105(3):968–75. PMID: [4926688](https://pubmed.ncbi.nlm.nih.gov/4926688/)
56. Goswami M, Mangoli SH, Jawali N. Involvement of reactive oxygen species in the action of ciprofloxacin against *Escherichia coli*. *Antimicrob Agents Chemother.* 2006; 50(3):949–54. PMID: [16495256](https://pubmed.ncbi.nlm.nih.gov/16495256/)
57. Kaldalu N, Mei R, Lewis K. Killing by ampicillin and ofloxacin induces overlapping changes in *Escherichia coli* transcription profile. *Antimicrob Agents Chemother.* 2004; 48(3):890–6. PMID: [14982780](https://pubmed.ncbi.nlm.nih.gov/14982780/)
58. Nishino K, Yamaguchi A. Analysis of a complete library of putative drug transporter genes in *Escherichia coli*. *J Bacteriol.* 2001; 183(20):5803–12. PMID: [11566977](https://pubmed.ncbi.nlm.nih.gov/11566977/)
59. Lawhorn BG, Gerdes SY, Begley TP. A genetic screen for the identification of thiamin metabolic genes. *J Biol Chem.* 2004; 279(42):43555–9. PMID: [15292217](https://pubmed.ncbi.nlm.nih.gov/15292217/)



60. Spoering AL, Vulic M, Lewis K. GlpD and PlsB participate in persister cell formation in *Escherichia coli*. *J Bacteriol*. 2006; 188(14):5136–44. PMID: [16816185](#)
61. Doi Y, Arakawa Y. 16S ribosomal RNA methylation: emerging resistance mechanism against aminoglycosides. *Clin Infect Dis*. 2007; 45(1):88–94. PMID: [17554708](#)
62. Kwon DH, Dore MP, Kim JJ, Kato M, Lee M, Wu JY, et al. High-level beta-lactam resistance associated with acquired multidrug resistance in *Helicobacter pylori*. *Antimicrob Agents Chemother*. 2003; 47(7):2169–78. PMID: [12821464](#)
63. Webber MA, Piddock LJ. The importance of efflux pumps in bacterial antibiotic resistance. *J Antimicrob Chemother*. 2003; 51(1):9–11. PMID: [12493781](#)
64. Clarke KR. Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*. 1993; 18(1):117–43.
65. Warnecke TE, Lynch MD, Karimpour-Fard A, Lipscomb ML, Handke P, Mills T, et al. Rapid dissection of a complex phenotype through genomic-scale mapping of fitness altering genes. *Metab Eng*. 2010; 12(3):241–50. doi: [10.1016/j.ymben.2009.12.002](#) PMID: [20060059](#)
66. Gower JC. Generalized procrustes analysis. *Psychometrika*. 1975; 40(1):33–51.
67. Forsberg KJ, Patel S, Gibson MK, Lauber CL, Knight R, Fierer N, et al. Bacterial phylogeny structures soil resistomes across habitats. *Nature*. 2014; 509(7502):612–6. doi: [10.1038/nature13377](#) PMID: [24847883](#)
68. Sommer MO, Dantas G, Church GM. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science*. 2009; 325(5944):1128–31. doi: [10.1126/science.1176950](#) PMID: [19713526](#)
69. Robinson DG, Chen W, Storey JD, Gresham D. Design and analysis of Bar-seq experiments. *G3 (Bethesda)*. 2014; 4(1):11–8.
70. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26(1):139–40. doi: [10.1093/bioinformatics/btp616](#) PMID: [19910308](#)
71. Gilbert JA, Jansson JK, Knight R. The Earth Microbiome project: successes and aspirations. *BMC Biol*. 2014; 12:69. doi: [10.1186/s12915-014-0069-1](#) PMID: [25184604](#)