Markov Networks of Collateral Resistance: National Antimicrobial Resistance Monitoring System Surveillance Results from *Escherichia coli* Isolates, 2004-2012

William J. Love[1]*, Kelson A. Zawack[2,3], James G. Booth[2], Yrjo T. Gröhn[4], Cristina Lanzas[1]

1 Department of Population Health and Pathobiology, College of Veterinary Medicine, North Carolina State University, Raleigh, North Carolina, United States of America, **2** Department of Biological Statistics & Computational Biology, College of Agriculture and Life Sciences, Cornell University, Ithaca, New York, United States of America, **3** Tri-Institutional Program in Computational Biology & Medicine, New York City, New York, United States of America, **4** Department of Population Medicine and Diagnostic Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York, United States of America

* wjlove@ncsu.edu

Abstract

Surveillance of antimicrobial resistance (AMR) is an important component of public health. Antimicrobial drug use generates selective pressure that may lead to resistance against to the administered drug, and may also select for collateral resistances to other drugs. Analysis of AMR surveillance data has focused on resistance to individual drugs but joint distributions of resistance in bacterial populations are infrequently analyzed and reported. New methods are needed to characterize and communicate joint resistance distributions. Markov networks are a class of graphical models that define connections, or edges, between pairs of variables with non-zero partial correlations and are used here to describe AMR resistance relationships. The graphical least absolute shrinkage and selection operator is used to estimate sparse Markov networks from AMR surveillance data. The method is demonstrated using a subset of *Escherichia coli* isolates collected by the National Antimicrobial Resistance Monitoring System between 2004 and 2012 which included AMR results for 16 drugs from 14418 isolates. Of the 119 possible unique edges, 33 unique edges were identified at least once during the study period and graphical density ranged from 16.2% to 24.8%. Two frequent dense subgraphs were noted, one containing the five β-lactam drugs and the other containing both sulfonamides, three aminoglycosides, and tetracycline. Density did not appear to change over time (p = 0.71). Unweighted modularity did not appear to change over time (p = 0.18), but a significant decreasing trend was noted in the modularity of the weighted networks (p < 0.005) indicating relationships between drugs of different classes tended to increase in strength and frequency over time compared to relationships between drugs of the same class. The current method provides a novel method to study the joint resistance distribution, but additional work is required to unite the underlying biological

and genetic characteristics of the isolates with the current results derived from phenotypic data.

Author Summary

Surveillance of antimicrobial resistance patterns is an important responsibility in modern public health. Due to the genetic configuration of bacterial pathogens, the use of a single antimicrobial drug may select for bacteria that are resistant to multiple other antimicrobial drugs via set of mechanisms collectively known as "collateral resistance". We have developed a new analytic method to study and describe possible collateral resistance pathways using existing antimicrobial resistance surveillance data. The method, named "R-nets", use network models to visualize the patterns of collateral resistance which allows large amounts of information to be clearly communicated to public health officials and researchers. Applying the R-nets to publicly available data from *E. coli* collected by the FDA and USDA between 2004 and 20012 we found that the number of collateral resistance links was relatively constant, but there may be an increase in collateral resistances between drugs of different structures. The method described will hopefully be useful for predicting and managing the proliferation of highly resistant bacterial pathogens

Introduction

The evolution of acquired antimicrobial resistance (AMR) in pathogenic microorganisms is one of the foremost challenges in public health today. Antimicrobial drug use in medicine and agriculture generates selective pressure that selects for AMR in bacterial populations and facilitates emergence of multiple drug resistant (MDR) phenotypes [1]. Bacterial pathogens with an MDR phenotype pose a substantial clinical challenge since the antimicrobial drugs typically prescribed may not effectively clear a patient's infection and delay the patient's recovery. The most recent and dramatic example is the emergence of plasmid-mediated resistance to colistin in *Escherichia coli* isolated from animals and humans [2]. Colistin is the last resort to treat some infections caused by Gram-negative bacteria such as carbapenem resistant *Actinobacter baumannii* and at least one pan-resistant strain of *E. coli* has been isolated from a clinically ill human patient [3, 4].

Multiple mechanisms of bacterial evolution, including mutation, recombination, and clonal expansion, give rise to highly resistant strains of bacteria and encourage persistence of these strains following their emergence [5]. Genetic capitalism describes the phenomenon by which the progeny of a microbe with one fitness trait, a drug-resistant phenotype for example, tend to survive serial selection events, in turn increasing the likelihood that the progeny will acquire additional fitness traits via recombination [6]. The initial fitness trait may be acquired by mutation, horizontal gene transfer (HGT), or novel recombination event with other bacteria in the environment. An example of genetic capitalism was the rapid emergence and expansion of a fluoroquinolone-resistant strain of methicillin-resistant *Staphylococcus aureus* (MRSA) at the Atlanta Veterans Affairs Medical Center, where the proportion of fluoroquinolone-resistant MRSA isolates increased from 0 to nearly 80% within 1 year of the introduction of ciprofloxacin to the hospital's formulary [7]. Collateral selection, another mechanism capable of generating MDR strains, describes the phenomenon where selection pressure from one antimicrobial drug may additionally select for or against phenotypic resistances to other drugs via several

mechanisms [8]. Cross-resistance describes resistance to several related drugs by a single mechanism, e.g., point mutations to the DNA gyrase subunit A gene (*gyrA*) that increase resistance to quinolone antibiotics [9], co-resistance describes a set of resistances conferred by a set of frequently concurrent genes, e.g., polymixin and colistin resistance genes carried on the recently-described MCR-1 plasmid in *Escherichia coli* [2], and pleiotropic resistance describes individual mutations that simultaneously affect resistances for multiple unrelated drugs such as *marR* gene alterations in *E. coli* that increase resistance to tetracycline, chloramphenicol, and fluroquinolones drugs [10, 11].

Monitoring the proliferation of existing MDR strains and emergence of novel MDR strains are primary goals of AMR surveillance programs carried out by governmental agencies across the world including, but not limited, to the Department of Agriculture (USDA), Food and Drug Administration (FDA) and Centers for Disease Control and Prevention (CDC) in the United States [12–14], the Department of Health in the United Kingdom [15], the European Centers for Disease Control and Prevention (ECDC) in the European Union [16, 17], and the collaborative Transatlantic Task Force on Antimicrobial Resistance created in 2009 [18, 19]. These agencies' survey methodologies and their published reports have been largely focused on univariate phenotypic resistances and the prevalence of MDR strains [20]. These surveillance reports however provide little information about the joint distributions of drug resistances in the overall population that may contribute to MDR strain development and emergence via collateral resistance. Outside of AMR surveillance, several studies of *in vitro* and *in situ* bacterial populations have begun to explore joint resistance distributions in an effort to identify strategies to mitigate MDR development [21–23].

Studying joint distributions of resistances in AMR surveillance data is necessary to understand *in situ* MDR strain evolution, but poses a number of challenges. The number of estimated correlation coefficients required to fully describe the pair-wise resistance distributions grows quadratically with the number of drugs in a panel, specifically $_kC_2 = k^2/2 - k/2$ for a panel of k drugs; Most AMR panels contain a dozen drugs or more, requiring at least $_{12}C_2 = 66$ correlations to be estimated. Hypothesis testing may be used to determine which drug pairs are not correlated, but the large number of estimated correlations results substantial type I error rate inflation. Pair-wise correlations also do not control for confounding by other variables in the dataset.

Better methods are needed to characterize and quantify the joint resistance distributions in bacteria. We propose graphical models, specifically Markov networks, to study joint distributions in AMR surveillance. Graphical models are mathematical constructs which represent the interactions between elements in complex systems. A variety of well-studied parameters, e.g., density and modularity, have been described to summarize these models at multiple levels of complexity. This study's objective is to estimate the Markov networks' structures that represent existing AMR surveillance data to effectively describe and visualize AMR relationships among a population of bacteria. This technique is intended to supplement the current methods used to analyze and report AMR surveillance results. The method is demonstrated here using data collected by the National Antimicrobial Resistance Monitoring System for Enteric Bacteria (NARMS).

Materials and Methods

Surveillance data

The NARMS study tracks AMR in *E. coli*, *Salmonella* spp., *Enterococcus* spp., and *Campylobacter* spp. isolated from slaughter houses, retail meat, and cases of food borne illness by the USDA, FDA and CDC respectively. Surveillance results from NARMS between 1998 and 2013

were recently made publically available [24]. Isolates were tested for AMR with genus-specific drug panels, typically containing 13 to 15 drugs with resistance results reported as minimum inhibitory concentrations (MIC).

To demonstrate our approach, a subset of the NARMS data was selected from the AMR results from *E. coli* isolated between 2004 and 2012 ($n = 14,418$). The MIC results for the following 16 of the available 23 drugs were used to demonstrate the networks: ampicillin (AMP), amoxicillin and clavulanic acid (AMC), ceftriaxone (AXO), cefoxitime (FOX), ceftiofur (TIO), amikacin (AMI), gentamicin (GEN), kanamycin (KAN), streptomycin (STR), nalidixic acid (NAL), ciprofloxacin (CIP), sulfisoxazole (FIS), trimethoprim and sulfamethoxazole (COT), chloramphenicol (CHL), tetracycline (TET), and azithromycin (AZI). The drug resistances were grouped into classes based on the structure of the drug tested (Table 1). The breakpoints for these antimicrobials and summary of observed resistances in the NARMS data are provided in Table 2.

Additional information regarding the data and sample sizes is available in supplemental material (S1 Text).

Statistical analysis

All statistical analyses were performed using R version 3.2.3 [25]. Spearman's rank correlations were used to estimate MIC relationships and trends in graphical parameters over time [26]. Sparse graphical model structures were constructed using the glasso package version 1.8 [27] and graphical parameters were estimated using the igraph package version 1.0.1 [28].

Table 1. Descriptions of 16 antimicrobial drugs included in the AMR panel for *E. coli* isolates from chicken carcass rinsates and commercial chicken breasts between 2004 and 2012 collected by the NARMS study.

Drug	Abbrev.	Class	Target Mechanism (Targets)
Ampicillin	AMP	β-lactams	Cell Wall Synthesis (Penicillin Binding Protein)
Amoxicillin + Clavulanic Acid	AMC		
Ceftriaxone	AXO		
Cefoxitime	FOX		
Ceftiofur	TIO		
Amikacin*	AMI	Aminoglycosides	Protein Synthesis (30S Ribosomal Subunit)
Gentamicin	GEN		
Kanamycin	KAN		
Streptomycin	STR		
Ciprofloxacin	CIP	Quinolones	DNA Synthesis Inhibition
Nalidixic Acid	NAL		(Topoisomerase IV & Gyrase)
Trimethoprim + Sulfamethoxazole	COT	Sulfonamides	Folic Acid Synthesis (Dihydropteroate Synthase)
Sulfisoxazole	FIS		
Tetracycline	TET	Tetracyclines	Protein Synthesis
			(30S Ribosomal Subunit)
Chloramphenicol	CHL	Phenicols	Protein Synthesis
			(50S Ribosomal Subunit)
Azithromycin†	AZI	Macrolides	Protein Synthesis
			(50S Ribosomal Subunit)

*Amikicin was only included in the drug panels during 2004–10, and removed afterward.

†Azithromycin was included in panels during 2011 & 2012

doi:10.1371/journal.pcbi.1005160.t001

Table 2. Summary of AMR testing ranges, resistance breakpoints, and results from 14,418 *E. coli* isolates collected from chicken carcass rinsates and commercial chicken breasts between 2004 and 2012 to an AMR panel of 16 drugs by the NARMS study. The breakpoints were collected from CLSI and USDA-ARS documents [57, 63].

Drug	Min. Obs MIC*	Max. Obs MIC	Dil'ns Obs	Range Tested(# Dil'ns)	MIC Breakpt.	Min Res. (%)	Max Res. (%)	Mean % Res.
AMP	≤ 1	> 32	7	1 - 32 (6)	≥ 32	17.2	24.3	20.5
AMC	≤ 1	> 32	7	1 - 32 (6)	≥ 32*	5.6	14.9	10.7
AXO	≤ 1/4	> 64	10	1/4 - 64 (9)	≥ 4	4.4	13.4	9.8
FOX	≤ 1/2	> 32	8	1/2 - 32 (7)	≥ 32	5.0	14.1	10.3
TIO	≤ 1/8	> 8	8	1/8 - 8 (7)	≥ 8	4.4	10.5	7.9
AMI	≤ 1/2	≤ 16	6	1/2 - 64 (8)	≥ 64	0.0	0.0	0.0
GEN	≤ 1/4	> 16	8	1/4 - 16 (7)	≥ 16	30.8	42.0	38.0
KAN	≤ 8	> 64	5	8 - 64 (4)	≥ 64	5.7	10.6	8.4
STR	≤ 32	> 64	3	32 - 64 (2)	≥ 64	38.9	62.7	49.0
CIP	≤ 1/64	> 4	10	1/64 - 4 (9)	≥ 1†	0.1	0.6	0.3
NAL	≤ 1/2	> 32	8	1/2 - 32 (7)	≥ 32	2.3	7.4	4.2
COT	≤ 1/8	> 4	6	1/8 - 4 (6)	≥ 4*	3.1	9.9	6.7
FIS	≤ 16	> 256	6	16 - 256 (5)	≥ 512	39.2	51.4	47.5
TET	≤ 4	> 32	5	4 - 32 (4)	≥ 16	40.2	49.8	45.5
CHL	≤ 2	> 32	6	2 - 32 (5)	≥ 32	0.9	2.2	1.4
AZI	≤ 1/2	> 16	7	1/8 - 16 (8)	≥ 32	0.0	0.4	0.2

*Reported breakpoint for primary drug in combination: Amoxicillin in AMC and Sulfamethoxazole in COT

†Reported breakpoints conflict between those reported by CLSI (≥ 4) and the USDA-ARS website (≥ 1). The USDA-ARS breakpoint was used to define in the current study.

doi:10.1371/journal.pcbi.1005160.t002

Overview of network inference

A graphical model *G* of a system is comprised two sets: the vertex set *G(V)* which defines the system's *k* discrete elements and the edge set *G(E)* which defines the *m* pair-wise interactions between the system's elements. For simple graphical models, *G(E)* may be expressed as the set of unique, unordered pairs of adjacent vertices ($G(E) = \{(v_i, v_j)| v_i$ adjacent to $v_j, i \neq j\}$) or as an adjacency matrix [29].

A Markov network is a specific type of undirected graphical model used to represent relationships between variables in a data set [30, 31]. Each of the dataset's *k* constituent random variables are represented by a vertex in the Markov network ($G(V) = \{v_i, \ldots, v_k\}$). Edges in a Markov network are defined by the variables' partial correlations (ω_{ij}), which represent the standardized covariance of a pair of variables v_i and v_j when conditioned on all other variables in the dataset. When $\omega_{ij} = 0$, v_i and v_j are conditionally independent and are not adjacent in a Markov network. Therefore, a Markov network's edge list consists of the *m* unique, unordered pairs of variables which are not conditionally independent ($G(E) = \{(v_i, v_j)| \omega_{ij} \neq 0\}$) and the network's adjacency matrix **A** may be defined from **Ω** using the indicator function $\mathbb{1}_M$ (Eq 1).

$$A_{ij} = \mathbb{1}_M(\omega_{ij}) = [\omega_{ij} \neq 0] \qquad \text{Eq 1}$$

When the structure of a system's Markov network is unknown, it is possible to estimate the partial correlations and network structure from observed data. An empirical correlation matrix Σ can be inverted to produce an precision matrix **Θ** (**Θ** = Σ^{-1}), which in turn can be used to estimate **Ω** (Eq 2) [32].

$$\omega_{ij} = \frac{-\vartheta_{ij}}{\sqrt{\vartheta_{ii}\vartheta_{jj}}}$$ Eq 2

Graphical models estimated in this way are typically complete ($m = {}_kC_2$) since trivial correlations will persist, even after conditioning on the other variables in the set. However, graphical models, including Markov networks, are easier to interpret and are more useful when they are sparse ($m << {}_kC_2$) [31].

The least absolute shrinkage and selection operator (LASSO) is a method used to generate more parsimonious models of joint distributions. The graphical LASSO applies an L_1-regularization penalty ρ to estimate a penalized precision matrix (Θ^*) [27]. The penalization reduces the absolute value of all elements of Θ ($|\vartheta_{ij}| > |\vartheta^*_{ij}|$) and if the penalty is large enough, some ϑ_{ij}^* are reduced to zero. When estimated from Θ^* instead of Θ, the trivially small partial correlations are reduced to zero and the Markov network will be sparse. While any non-negative value may be assigned to ρ, its useful range is limited to $min|\vartheta_{ij}| < \rho < max(|\vartheta_{ij}|)$. When $\rho > max(|\vartheta_{ij}|)$, Θ^* and Ω are diagonal, all ω_{ij} are trivial, and $G(E) = \emptyset$; conversely, if $\rho < min(|\vartheta_{ij}|)$, then no ω_{ij} are trivial, and the resultant graphical model will typically be complete, similar to using the unpenalized precision matrix.

Graphical models of collateral resistance

The Markov networks of collateral resistance developed here are subsequently referred to as "R-nets" and two versions are presented: simple R-nets (R) and weighted R-nets (R'). For these networks, $R(V) = R'(V)$ and $R(E) = R'(E)$, but their adjacency matrices are different. For R, the simple adjacency matrix **A** is defined by Eq 1. For R', Ω is used as the weighted adjacency matrix where the elements ω_{ij} represent the strength and direction of the relationship. The vertices of an R-net represent the observed distribution antimicrobial resistances, e.g., TIO represents the observed set of MIC_{TIO}. The edges of an R-net represent correlated resistances, which in turn identify potential collateral resistances in the population. A Spearman's rank correlation matrix Σ_y was estimated for the adjusted log_2MIC values within each year y. A non-parametric estimator was used because the distribution of the MICs, even when transformed, frequently did not conform to a normal distribution, and Pearson's correlation estimates may yield biased results when applied to non-normal data [26]. The graphical LASSO and Eq 2 were used to estimate the structures of R_y and R'_y. The graphical LASSO method assumes multivariate normal data, but it has been demonstrated that an untransformed Spearman's correlation matrix may be substituted for the more traditional parametric correlation matrix, provided the former is positive semidefinite [33].

Regularization penalty selection

The structures of sparse Markov networks generated via the graphical LASSO are conditional on the L_1-penalty used to estimate Θ^*, thus the selection of the penalty is an important step in the inference process. In general, the value of ρ should be low enough to assure no important edges are lost, and high enough to effectively reduce the density of the graph. Comparisons of R_y and R'_y over time were restricted to sets of R-nets generated from a common penalty to improve comparability of the networks. To select a common penalty value, 12 sets of R and R' were produced and evaluated with each set of networks generated using a separate value of ρ between 0.05 and 0.60. The structure and distribution of density ($\overline{m}_y$; Eq 3) within each set of

R-nets were subjectively evaluated to select a single penalty value to apply across all years. The common penalty was selected via a supervised review of the R-nets based on network interpretability and trends in $\overline{m}_y$ over ρ. A network should be sparse enough to be understood when visualized, meaning the graph should not be too dense for a viewer to understand the set of edges present. Edges representing expected cross-resistances, e.g., NAL-CIP, AMP-AMC, and FIS-COT, should be strong and frequently present. Additionally, the number of unstable cycles (cycles with an odd number of negative edges) should be minimized [34]. The median density ($\overline{m}_{50}$) and range of $\overline{m}_y$ was plotted against the evaluated values of ρ and a point was sought where the slope of the line broke from steeply decreasing to level off; this method is similar to the Scree test for factor retention in FA and PCA [35, 36]. The lowest penalty that generated R and R' that met these criteria were selected for further evaluation.

$$\overline{m}_y = \frac{m_y}{{}_kC_2}$$
Eq 3

Longitudinal comparison of graphical parameters

Three graphical parameters were used to evaluate changes in AMR resistance over time described by R and R': density, vertex degree, and modularity. Density quantifies a graph's overall interconnectivity and trends in $\overline{m}$ could indicate overall trends in AMR relationships and risk of collateral resistance in the population over time [37]. Increasing density over time represents more interconnectivity of drug resistances.

Vertex degree d_i is equal to the number of other vertices adjacent to v_i. In an R-net, d describes the number of other resistances one MIC is related to in the population. In the R-nets vertices with high d_i may represent resistances which are influential to the development of MDR strains. For example, if resistance to drug A had a high degree in a bacterial population, e.g., $d_A = 5$, the use of drug A could select for increased resistance to A and potentially affect resistance to five other drugs, assuming that the covariance structure of phenotypic drug resistance identified by the Markov networks reflects genetic mechanisms allowing for collateral resistance. The high vertex degree of A does not indicate that the drug resistance is responsible for resistance noted in five adjacent vertices, only that selection for A could possibly influence the other resistances.

Modularity (Q) measures how frequently adjacent vertices are similar or dissimilar as defined by vertex attributes [38]. The class of drug associated with the resistance of each vertex was used to assign group membership (Table 1). Modularity is positive when edges join similar vertices more frequently than would be expected by chance and negative when dissimilar vertices are more frequently than would be expected by chance. Trends in modularity may represent shifting tendencies in resistance relationships to exist between more similar or less similar drugs. Similar estimates of modularity for signed and weighted networks (Q') can be estimated [39]. The value of Q' describes the relative strength of edges between similar vertices compared to dissimilar vertices. Spearman's rank correlation was used to test for a trends in $\overline{m}$, Q, and Q' over time due to the small number of values being compared [26].

Principal component analysis

Principal component analysis has been used to study resistance relationships and was performed here to evaluate how the current method compares to a previously employed method [23]. Three years of the study were selected to represent data from the beginning, middle, and end of the study, respectively, and PCA was performed separately on the data from each year. The eigenvalues of Σ_y (λ) were computed and the components for which $\lambda > 1$ were extracted and subsequently oblimin rotated [40, 41]. The oblimin rotation, an

oblique rotation method, was used to allow the loadings to assort naturally without imposing orthogonality. Drug resistances with component loadings that had an absolute value greater than 0.4 were considered to be important and were assigned to the respective rotated component [42]. The rotated components were compared to the respective R-nets to evaluate agreement between the two methods.

Results

Regularization penalty selection

The range of ρ evaluated was fully contained in the useful range identified above, with $0 < |\vartheta_{ij}| < 2.7$ for all years in the study. The upper region of the possible penalties ($0.6 < \rho < 2.7$) was not evaluated since the networks generated in this region would have produced networks too sparse to be informative. The density of R-nets over all ρ values and years ranged from 4.8% to 57.1% over the evaluated range of ρ, and $0.3 \leq \rho \leq 0.5$ generated graphs with very similar densities between 10% and 20%, indicating consistent graph structures in this range of ρ (Fig 1) and changes in the slope of $\overline{m}_{50}$ over ρ were noted at $\rho = 0.10$ and $\rho = 0.25$ (Fig 2). In general, the R-nets generated by $\rho = 0.10$ were too dense to easily interpret and several unstable cycles were noted in R'_{2008}, R'_{2009} and R'_{2010}. The R-nets generated by $\rho = 0.25$ were sparse enough to be interpreted with reasonable effort. Additionally, R' under $\rho = 0.25$ contained no unstable cycles since all partial correlations were positive. The latter penalty of $\rho = 0.25$ was selected as the common penalty and used to generate R and R' interpreted below (Fig 3). The supplemental material provides an overview of ρ's impact on density and modularity (S2 Text) and a more in-depth description of R and R' conditioned on $\rho = 0.10$ (S3 Text).

Network structure

A total of 119 unique edges could have observed during the entire study period (the AZI-AMI edge could not have been observed since both drugs were never included in the same year), and 105 edges could have been observed in each year. Of these 119 unique edges, 33 unique edges were observed in at least one year of the study, and 15 appeared in all 9 years. A bimodal distribution was noted in the frequency of edge appearance, with one group of edges observed in 1 to 4 years, and another group of edges observed for 6 years or more (Fig 4). Sixteen of the 33 observed edges found represented resistance relationships between drugs of the same class (blue areas of Fig 4), most of which were present in all nine years during the study (13/16 within-class edges) indicating the relative stability of these relationships during the study period (Fig 5A). Seventeen between-class edges (red areas of Fig 4) were identified, of which only 4 were present for 8 or 9 years. The majority of between-class edges (12/17 between-class edges) were present for 3 years or less (Fig 5B).

Graphical parameters

No significant trend in $\overline{m}$ over time was noted (Spearman's rho = -0.14, p = 0.71). Two dense subgraphs of R were noted to be frequently present. The first frequent dense subgraph R_{BLA} contained the five β-lactam drug resistances AMC, AMP, AXO, FOX, and TIO and formed a clique in every R_y. The second frequent dense subgraph R_{AST} included GEN, KAN, STR, COT, FIS and TET, and $\overline{m}_{AST}$ exceeded 50% in six out of the nine years.

Individual vertices demonstrated several different patterns of degrees during the study period, and $d \leq 6$ for all vertices and time periods. For the β-lactam drugs, $d = 4$ was typical but occasionally increased to 5 or 6. Different patterns were noted among the degrees of the aminoglycosides: d_{KAN} was high the earlier in the study and decreased in later years ($3 \geq d_{KAN}$
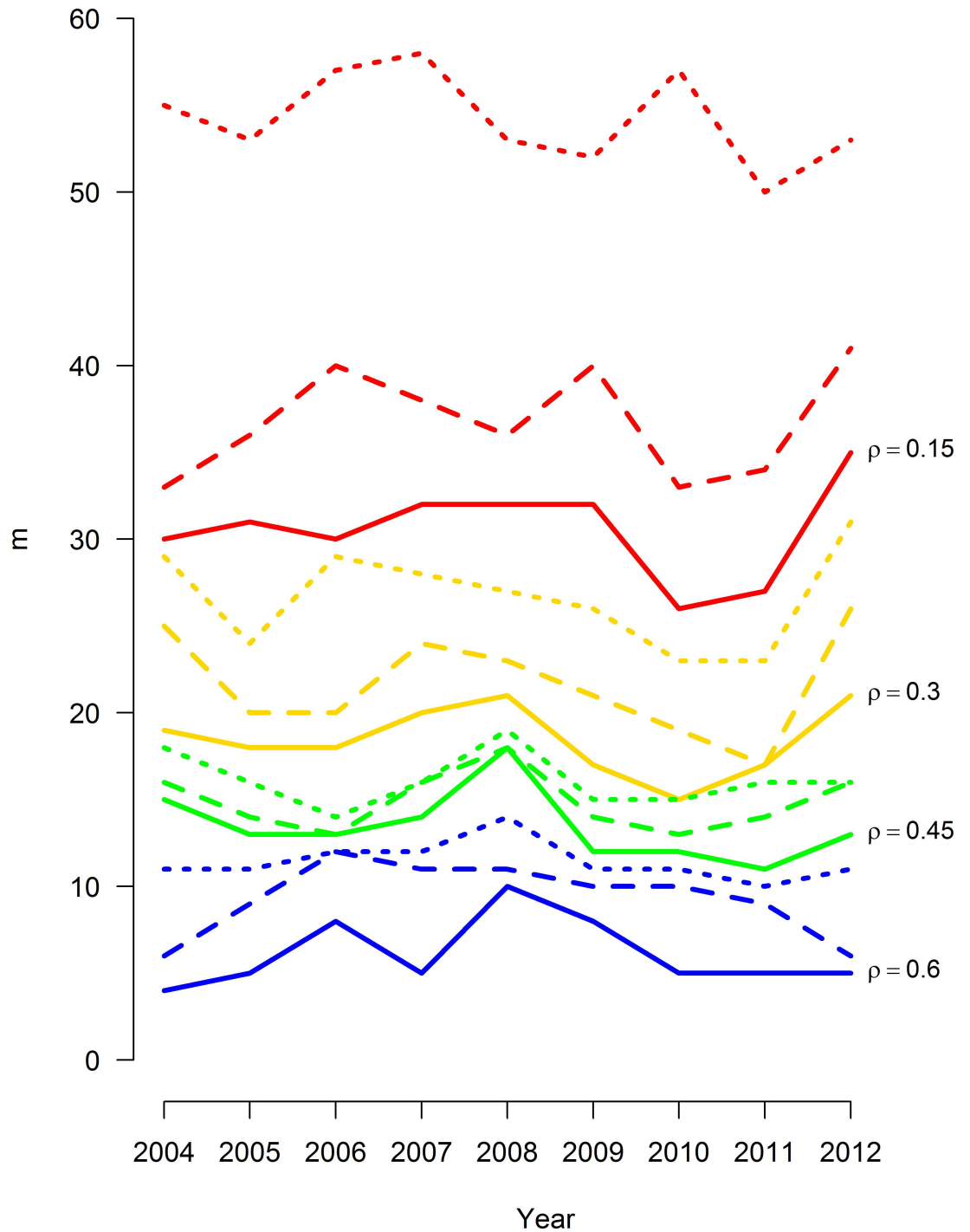
Fig 1. Size of Markov networks of AMR surveillance data over time, stratified by L₁ regularization value.
Networks were generated via the graphical least absolute shrinkage and selection operator with 12 L₁ penalty values
(ρ) from 0.05 to 0.60. Antimicrobial data includes 16 antimicrobial drugs results from 14,418 *E. coli* isolates collected
from chicken carcass rinsates and commercial chicken breasts between 2004 and 2012.

in 2009 and prior and $2 \leq d_{\mathrm{KAN}}$ in 2010 and later), d_{GEN} and d_{AMI} were more consistent, and
d_{STR} varied widely and without a clear pattern. In every time period, $d_{\mathrm{CIP}} = 1$, corresponding to
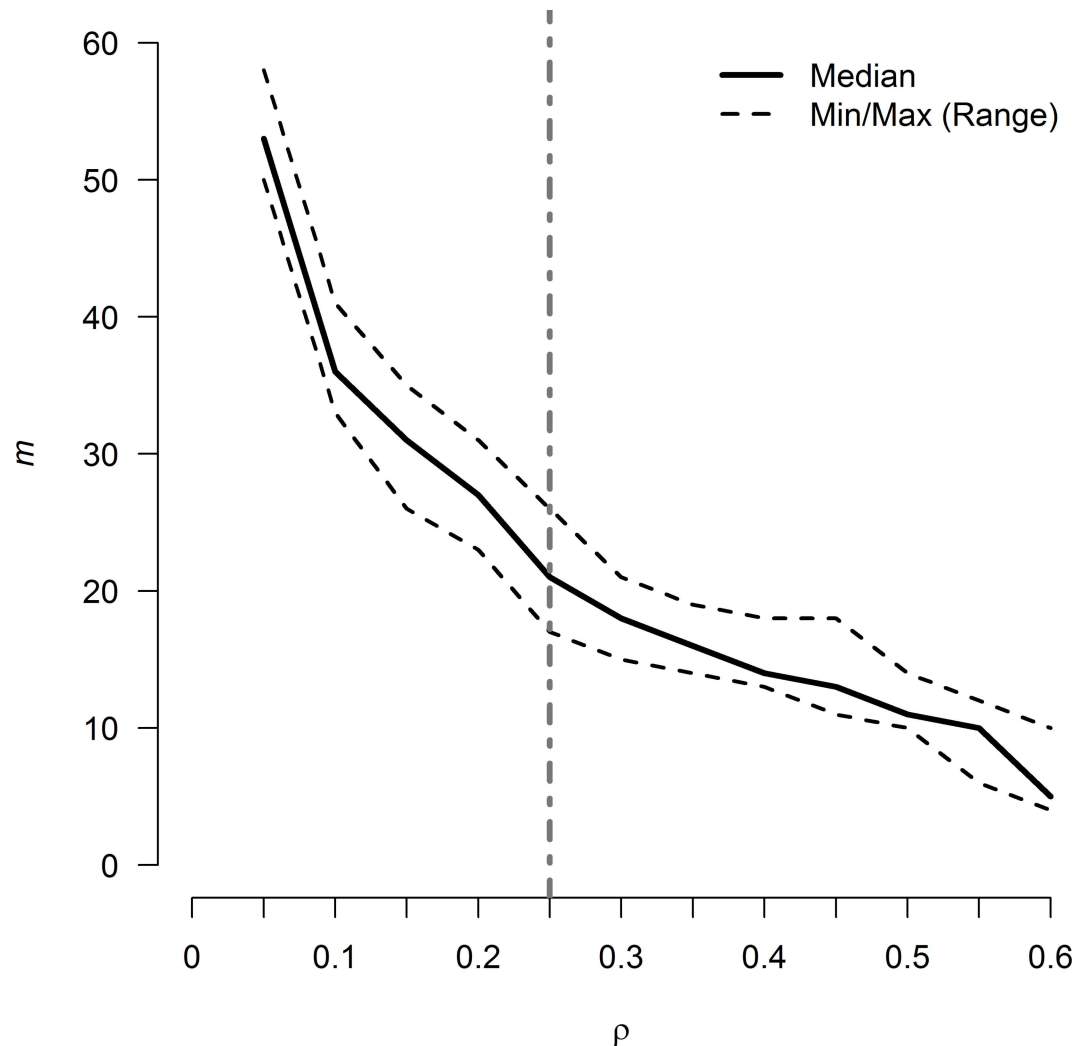
Fig 2. Size of Markov networks of AMR surveillance data (m) over L_1 regularization values. R-nets were generated using 12 L_1 penalty values (ρ), from 0.05 to 0.60 in increments of 0.05 for all 9 time periods in the study. The solid line represents the median graph size (m_{50}) for the set R, which contains one R-net for each of the 9 years in the study at each evaluated value of ρ, and the dashed lines above and below m_{50} represent the maximum and minimum size of R, respectively. R-net size decreased rapidly until $\rho \sim 0.3$, where the slope started to stabilize. The slope of m_{50} appeared to change most abruptly near $\rho = 0.10$ and $\rho = 0.25$. The vertical line at $\rho = 0.25$ represents the set of R-nets presented in Fig 3. The set of R-nets generated at $\rho = 0.10$ are presented in S3-1 Fig with further discussion in S3.

its edge with NAL, but d_{NAL} tended to increase through the study. The values for d_{CHL}, d_{FIS}, d_{COT}, and d_{AZI} changed little during the study period. There was substantial variation in d_{TET} and did not appear to follow a pattern.

Modularity defined by class in the unweighted R-nets ranged from $Q_{2012} = 0.267$ to $Q_{2005} = Q_{2006} = 0.420$. Modularity was not significantly associated with time (Spearman's rho = -0.49, p = 0.18; Fig 6). When weighted by partial correlation, modularity estimates ranged from $Q'_{2012} = 0.321$ to $Q'_{2005} = 0.466$, and a statistically-significant decreasing trend was noted in Q' (Spearman's rho = -0.95, p < 0.005).

LEGEND

% Resistant

XXX   < 10%
XXX   10 - 20%
XXX   20 - 30%
XXX   30 - 40%
XXX   > 40%

Correlation Magnitude

0 < $|\omega_{ij}|$ < 0.05
0.05 < $|\omega_{ij}|$ < 0.10
0.10 < $|\omega_{ij}|$ < 0.20
0.20 < $|\omega_{ij}|$

Correlation Sign
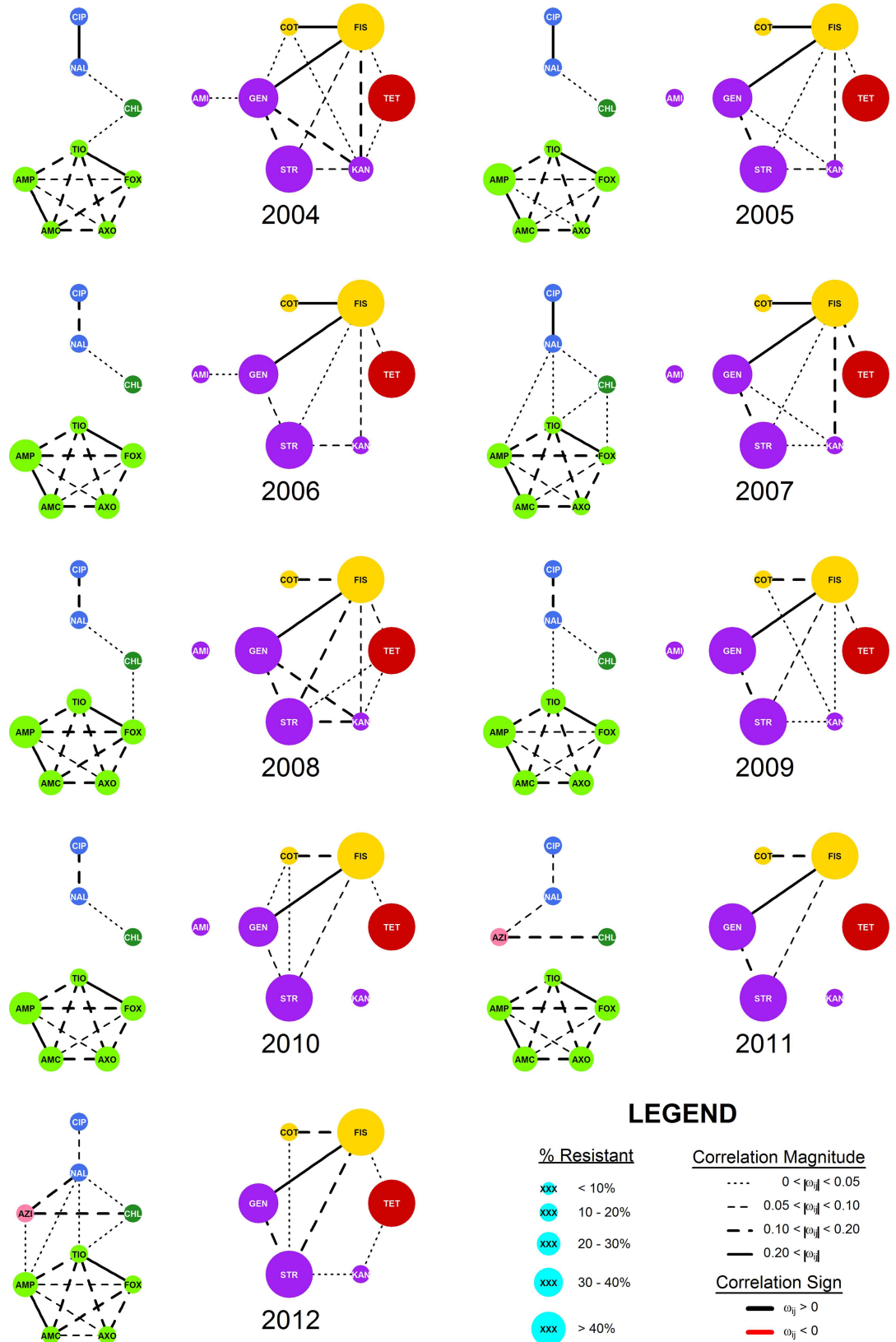
$\omega_{ij} > 0$
$\omega_{ij} < 0$

Fig 3. Markov networks of AMR surveillance data. Nine weighted Markov networks representing AMR data consisting of 14,418 total isolates of *E. coli* collected from chicken carcass rinsates and commercial chicken breasts from each year from 2004 to 2012. Networks were generated via the graphical least absolute shrinkage and selection operator with $\rho = 0.25$. The edges were decorated with line weights and styles to indicate the relative $|\omega_{ij}|$ magnitude. Vertex size indicates the percent of isolates with an MIC meeting or exceeding the published breakpoint for respective drug (Table 2). Vertex colors indicate the Class of drug associated resistance as follows: Vertex colors indicate the Class of drug associated resistance as follows: β-lactams–light green, quinolones–blue, aminoglycosides–purple, sulfonamides–yellow; chloramphenicol–dark green; tetracycline–red; macrolide–pink.

doi:10.1371/journal.pcbi.1005160.g003
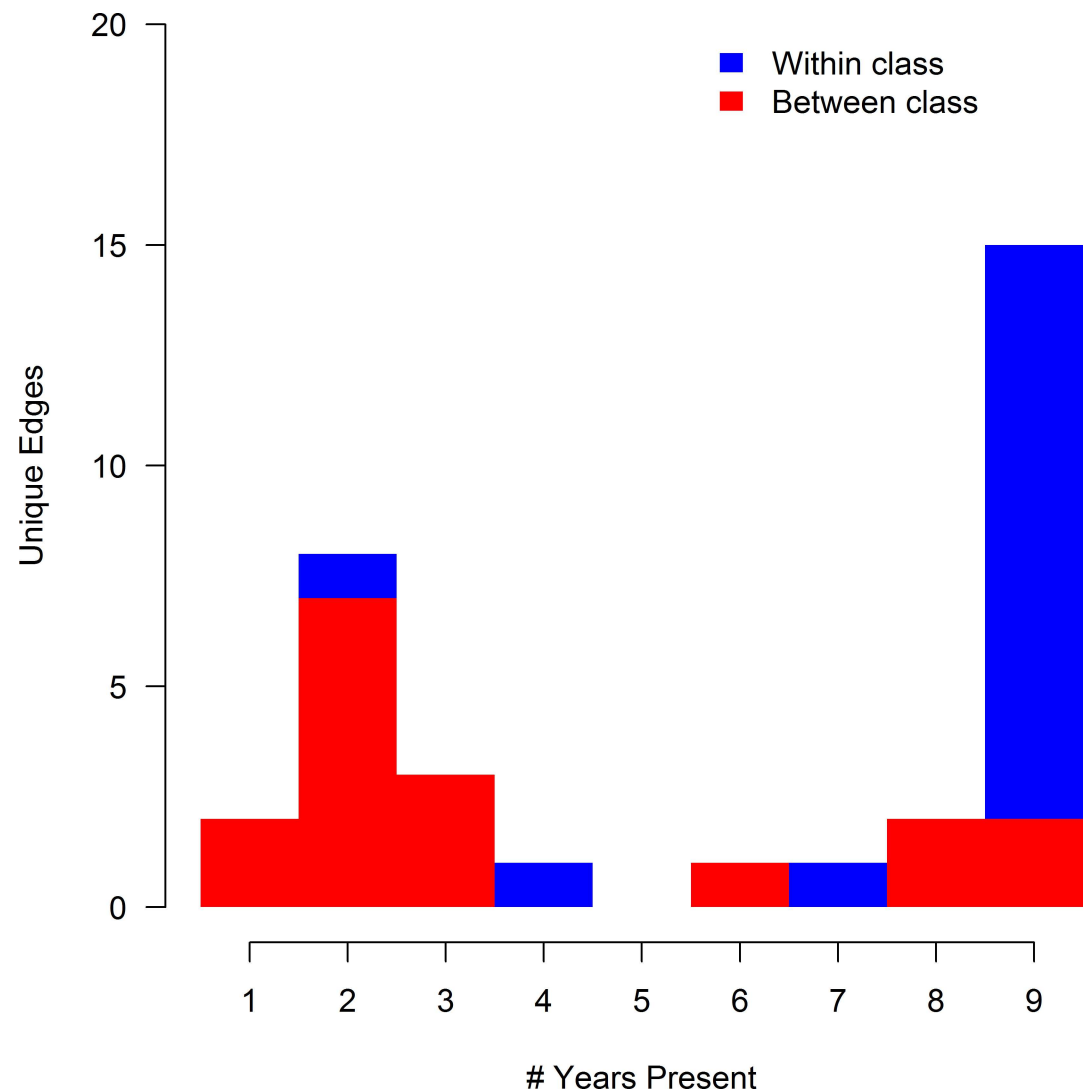
Fig 4. Histogram of edge frequency in Markov networks of AMR surveillance data. Thirty-three unique edges were identified in 9 Markov networks estimated using AMR data from 14,418 total isolates of *E. coli* collected from chicken carcass rinsates (USDA-ARS) and commercial chicken breasts (FDA) between 2004 and 2012. Networks were generated via the graphical least absolute shrinkage and selection operator with $\rho = 0.25$. Edges are categorized by the types of antibiotic resistances they joined, either resistances from the same class (blue) or different classes (red).
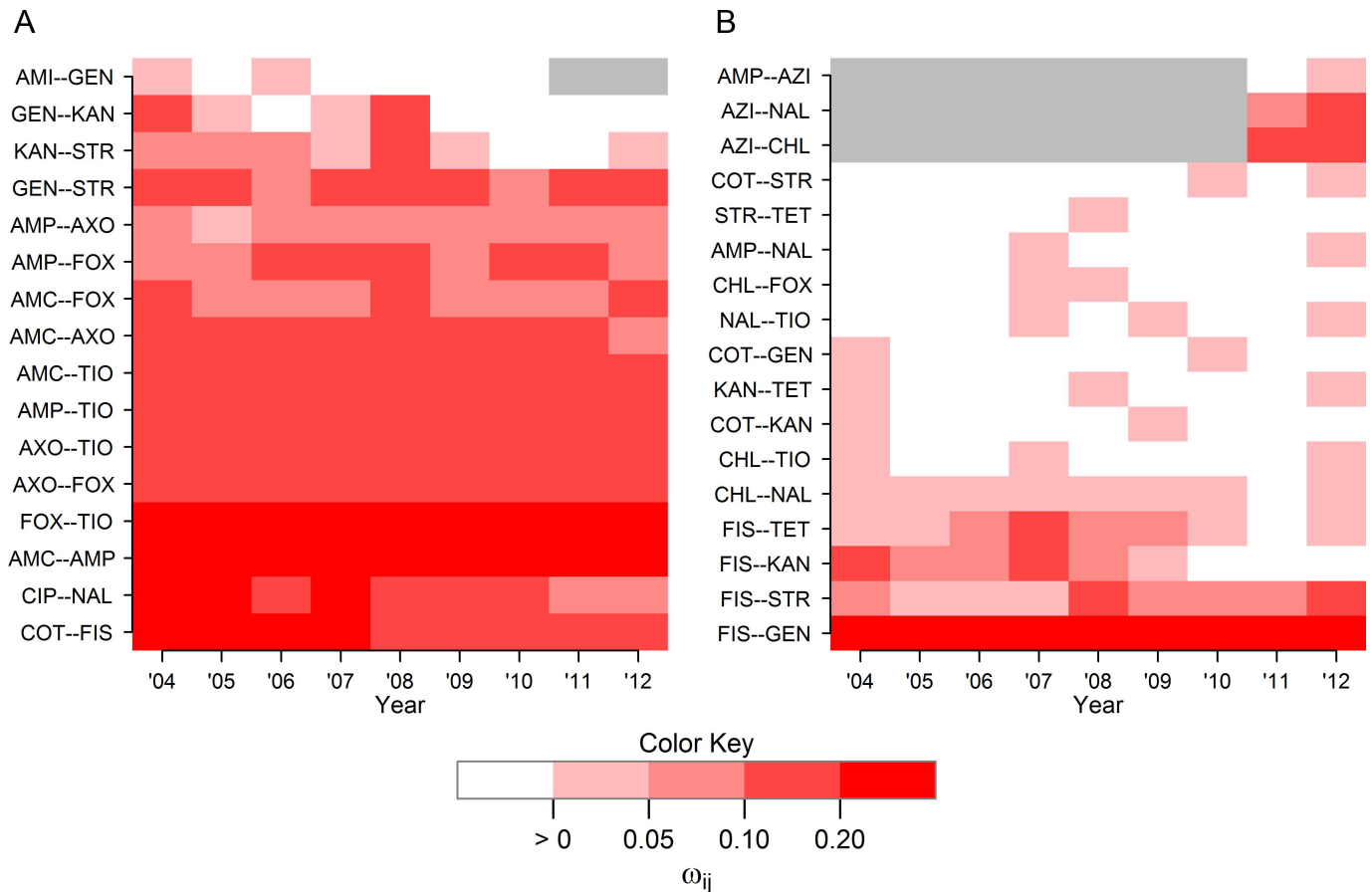
doi:10.1371/journal.pcbi.1005160.g004

Fig 5. Heatmap of non-zero partial correlations between drug resistances over time. (A) Edge weights for 16 edges between resistances of drugs of the same class and (B) 17 edges between resistances to drugs of different classes in Markov networks of AMR data. Resistance data was collected from 14,418 total isolates of *E. coli* collected from chicken carcass rinsates and commercial chicken breasts between 2004 to 2012. Networks were generated via the graphical least absolute shrinkage and selection operator with ρ = 0.25. Grey areas represent years of missing data due one of the antibiotic drugs was not included in the panel (AMI after 2010 and AZI prior to 2011).

doi:10.1371/journal.pcbi.1005160.g005

Principal component analysis

Principal component analysis was performed on Σ_{2004}, Σ_{2008}, Σ_{2012}. In each of the three years, four components were identified for extraction and oblimin rotation. Combined, these rotated components accounted for 61%, 64%, and 61% of the overall variance in $\log_2$MIC in 2004, 2008, and 2012 respectively. The loadings for each component were similar, though not identical, over all three years (Table 3). Correlations between the rotated components were small, with -0.10 ≤ r ≤ 0.25. Most rotated components aligned with subgraphs with $\overline{m}$ > 50% (Fig 7).

Discussion

The representation of AMR surveillance data using Markov networks generated via the graphical LASSO is a novel method to characterize potential collateral resistances in bacterial populations. The graphical nature of this method lends itself to simple visualization which allows complex relationships to be communicated clearly. The structures of R_{2004}, R_{2008}, and R_{2012} are similar to the variance structures identified by the respective PCA results, but the R-nets provide results that are simultaneously more detailed and more interpretable than the results from PCA.
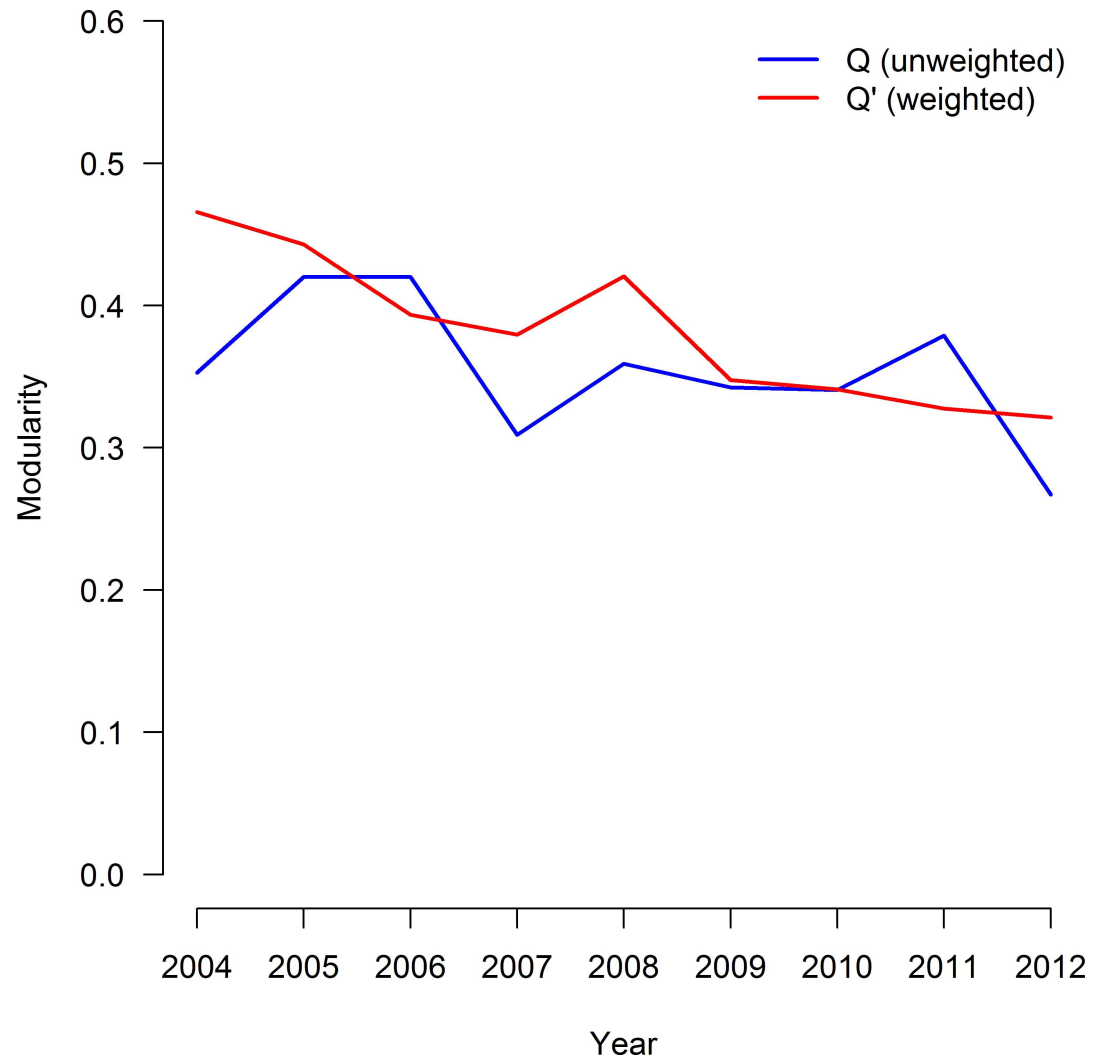
Fig 6. Modularity in Markov networks of AMR surveillance data over time. The unweighted modularity (Q, blue line) and weighted modularity (Q', red line) over time for Markov networks of AMR data based on the class (See Table 1). The data set consisted 14,418 total isolates of *E. coli* collected from chicken carcass rinsates and commercial chicken breasts from each year from 2004 and 2012. Networks were generated via the graphical least absolute shrinkage and selection operator with $\rho = 0.25$. A statistically negative trend over time was noted in Q' (p < 0.005), but not in Q (p = 0.18).

doi:10.1371/journal.pcbi.1005160.g006

Several resistance relationship patterns appeared in the models over time. Resistances to the β-lactam drugs were consistently and strongly related to each other in all years, generating the complete induced subgraph R_{BLA}. These patterns of related classes are likely an example of cross-resistance, though pleiotrophic and co-resistance mechanisms may also be present. The elements of $R_{\text{AST}}(V)$ did not represent resistances to a single class of drugs or even drugs targeting a common metabolic pathway. A similar grouping of resistances without GEN was previously described in beef cattle, where it was attributed to antibiotics frequently used in production medicine [23]. The quinolone drugs were always correlated with each other, and while CIP was only correlated with NAL, NAL was additionally correlated with CHL, TIO, and AMP in some years. Additionally, these patterns are consistent with those seen in the PCA of

Table 3. Eigenvalues (λ) and loadings for the oblimin-rotated components identified in the principal component analysis (PCA) of the Spearman's rank correlation matrix of MICs to 15 drugs from the NARMS study in 2004 (n_{2004} = 2097), 2008 (n_{2008} = 1292), and 2012 (n_{2012} = 1376). Variables with loadings > 0.4 were used to interpret the component, and the density of the induced subgraphs containing the variables with loading > 0.4 for each component are provided in $\overline{m}_\lambda$. The $\overline{m}_\lambda$ of components consisting of a single variable were undefined (N/D). Resistance results for azithromycin (AZI) was not available in 2004 and 2008, and resistance results for amikacin (AMI) was not available in 2012.

	2004				2008				2012			
	λ_1	λ_2	λ_3	λ_4	λ_1	λ_2	λ_3	λ_4	λ_1	λ_2	λ_3	λ_4
	3.31	*2.94*	*1.70*	*1.20*	*3.71*	*3.10*	*1.57*	*1.29*	*3.41*	*2.72*	*1.78*	*1.21*
AMC	0.85				0.89				0.83			
AMP	0.81				0.85				0.81			
AXO	0.77				0.84				0.82			
FOX	0.79				0.83				0.79			
TIO	0.81				0.85				0.81			
AMI				0.91				0.81	-	-	-	-
GEN		0.73				0.83				0.76		
KAN		0.67				0.69				0.47		
STR		0.63				0.74				0.75		
CIP			0.78				0.87					0.94
NAL		0.85					0.80				0.55	0.49
FIS		0.88				0.87				0.85		
COT		0.71				0.59				0.65		
TET		0.49				0.5				0.46		
CHL			0.49					0.61			0.75	
AZI	-	-	-	-	-	-	-	-			0.82	
$\overline{m}_\lambda$ (%)	100	73.3	66.7	N/D	100	66.7	100	0	100	53.3	100	100

doi:10.1371/journal.pcbi.1005160.t003

the same data, and similar to patterns seen in a previous study of AMR relationships in *E. coli* [23].

Evaluation of the graphical parameters provided additional insight into changes in the joint distributions of resistance over time. No temporal trend was apparent in $\overline{m}$ indicating that, on average, the amount of interconnectivity of AMR in this population of *E. coli* did not change substantially over time. The negative trend noted in Q' indicates a shift towards stronger relationships between drugs of different classes, weaker relationships between drugs of the same class, or both. Visual evaluation of Fig 5A and 5B suggests both processes may be occurring, with within-class edges becoming slightly weaker over time and the appearance of more, stronger between-class edges. Lower weighted modularity values over time were consistent across the common penalties and may indicate a concerning increase in co-resistance and pleiomorphic mutations in this population of *E. coli*. Neither the source of the decreasing trend in Q' or the behavior of any specific edge can be assigned without isolate covariate data, which were not available for data from the NARMS study. Quantification of specific network structures into simple numeric criteria is one of the major advantages of graphical models. While some information about a graph's structure is lost when the structure is condensed into a simple criterion, these criteria greatly facilitate the comparison of graphs. Parallel interpretation of multiple criteria, as is done here with Q, Q', and $\overline{m}$, can provide a more complete description.

It should be emphasized that the R-nets describe the joint relationships of resistances at the population level, hence little can be inferred about the genotype or the phenotypic resistance profile of an individual isolate from these results. At best, a probabilistic statement can be made about some MIC values given knowledge about the others. It is also not currently
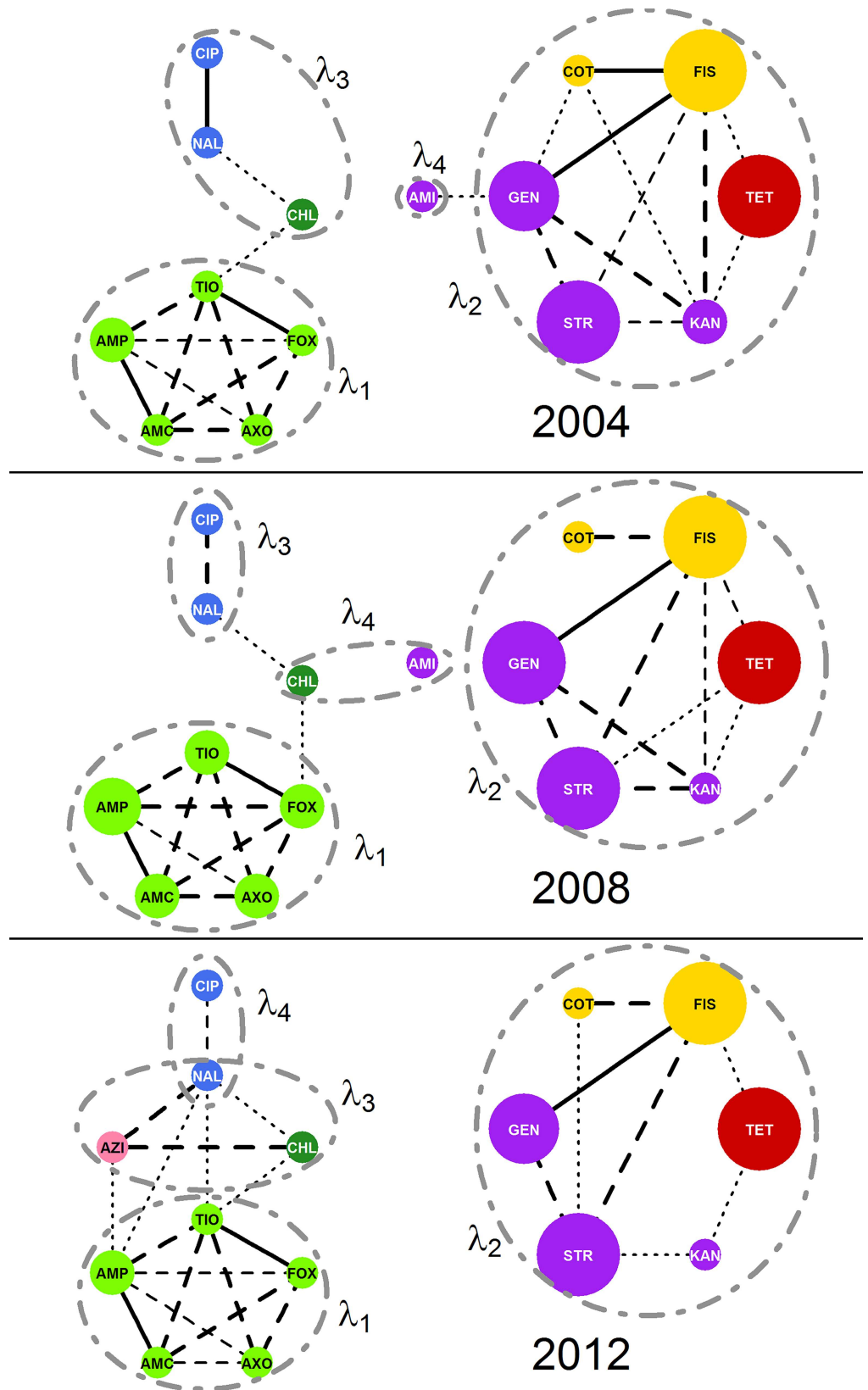
Fig 7. Markov network structure versus oblimin-rotated components. Comparison of Markov network structures (R'_y) and oblimin-rotated components (PCA) show good alignment indicating a both methods give similar result. Networks and components were estimated from *E. coli* isolates from chicken carcass rinsates and commercial chicken breasts collected in 2004 ($n_{2004} = 2097$), 2008 ($n_{2008} = 1292$; B), and 2012 ($n_{2012} = 1376$). Networks were generated via the graphical least absolute shrinkage and selection operator with $\rho = 0.25$. Sets of variables with loadings > 0.4 on each component are circumscribed with dashed circles and labeled with the component's eigenvalue (λ).

doi:10.1371/journal.pcbi.1005160.g007

possible to infer the source of the resistance or resistance relationships based on these models. The edges of the R-nets represent potential resistance relationships, but existence of the edge alone is not sufficient to induce collateral resistance. For example, even though GEN and FIS were adjacent, GEN will not affect FIS unless there exists a concurrent selection pressure for or against gentamicin resistance. Induction of collateral resistance for FIS by GEN requires the combination of the FIS-GEN resistance relationship and the selection for GEN, potentially from the therapeutic use of gentamicin. This illustrates that knowledge of the resistance relationships alone is insufficient to determine how the R-nets have influenced the patterns of resistance observed; antimicrobial use data are also needed and combining R-net results with exposure data is a topic of ongoing research. Despite this limitation, the R-nets can still identify phenotypic resistances of interest prospectively, specifically vertices of relatively high degree.

Application of the methods to multiple strains from varied sources will allow the opportunity to objectively identify resistance patterns, and potentially enable predictions about resistance relationship evolution based on previously observed R-nets. There are several applications where the population-level focus of the currently described method would be useful. First, this method could be directly applied to AMR surveillance in health care facilities where local evolution of highly resistant bacterial strains is a major concern [7]. While the dataset used to demonstrate the current method was relatively large and $n >> m$ even in the unpenalized case, the graphical LASSO can perform well even when the number of parameters to estimate exceeds the number of available observations [43]. Therefore, the current method may be employed to estimate R even when the sample size is modest, as may be the case in a single health care facility. R-nets can be estimated retrospectively from clinical data to monitor for antimicrobials with a high degree, indicating a high risk of extensive collateral resistance. If drug use data were available from patient records, limited inference could be made about how an MDR strain evolved in a clinical environment. If appropriate negative correlations were noted, a selection inversion could also be attempted [21]. The R-nets could also be applied to monitor emergence of novel resistance relationships at any scale over time. Knowledge of resistance relationship dynamics will help improve clinical decision making by informing the physician about what resistances may altered by use of a specific drug. The R-nets can help inform policy making by similarly tracking resistance relationships and detecting resistance of relationships of concern, and can help facilitate AMR research by screening large numbers correlations to locate non-trivial associations.

The R-nets revealed several interesting patterns of AMR involving the resistances the quinolones drugs NAL and CIP. In *E. coli*, quinolone resistance may be increased by mutations in the A and B subunits of DNA gyrase [9, 44] and mutations in *parC* and *parE*, which encode subunits of DNA topoisomerase IV[45–47]. Increases in NAL and CIP due to these mutations would be expected to lead to a strong relationship between NAL and CIP, but no other drugs in the panel since the quinolones are the only class of antimicrobials affected by DNA gyrase and topoisomerase mutations. This relationship was observed, with $\omega_{NAL,CIP} > 0.1$ in most years indicating a consistent strong NAL-CIP edge. Increased expression of active, non-specific efflux pumps in *E. coli* is an alternate and complementary mechanism of quinolone resistance,

and also increases resistance to chloramphenicol, tetracycline, and other classes of drugs [10, 11, 48]. If efflux pumps were an important mechanism of quinolone resistance in the sampled population of *E. coli*, correlations should be noted among NAL, CIP, TET and CHL and a dense subgraph including these vertices should be noted. However, no such subgraph was noted: TET was never found to be correlated with CIP, NAL, or CHL, and $\omega_{\text{CHL,NAL}} < 0.05$ most years. Efflux pump expression may explain the weak CHL-NAL edge but, overall, these results would appear to support the conclusion that efflux pumps expression was not an important source quinolone resistance, or at least was not a major contributor to resistance in this population.

Given $\rho = 0.25$, every edge in every year represented a positive partial correlation in resistances, indicating that, on average, increasing resistances to one drug was only associated with increased, and never decreased, resistances of the adjacent vertices. Among the unpenalized partial correlations, positive partial correlations outnumbered negative partial correlations by about a 2-to-1 margin, and the median absolute value of negative partial correlations was about half that of positive correlations. Negative partial correlations were only noted when $\rho \leq 0.15$. These findings are consistent with the phenomenon of genetic capitalism where the progeny of bacteria with at least one advantageous mutation tend to acquire other additional advantageous traits over time via recombination and HGT [6]. The relative weakness and infrequency of negative partial correlations compared to positive partial correlations is also consistent with the patterns seen in the resistance relationship networks of previous studies [22, 49]. One application of R-nets and similar networks is to identify pairs or larger groups of collaterally susceptible antibiotics to create a selection inversion: a reduction in overall AMR via strategic antibiotic use [21, 22]. Without any negative partial correlations, it is unlikely that a selection inversion could be achieved in this population of *E. coli*, but could be feasible in other populations.

The purpose of the L_1 penalization in the graphical LASSO procedure is to eliminate trivial edges from the graph by reducing their corresponding elements of Ω to zero, but the penalization also biases the non-trivial elements towards zero, as well [27, 50]. This is effect is the 'shrinkage' aspect of the least absolute shrinkage and selection operator. For example, the estimated unpenalized partial correlation between AMC and AMP in 2012 is 0.62, but when estimated via the graphical LASSO with $\rho = 0.25$ this estimated partial correlation is reduced to 0.36; the latter estimate of $\omega_{\text{AMC, AMP}}$ has been moved towards 0, or shrunken, compared to the former due to the L_1 penalization. This bias caused by the penalization should be kept in mind when interpreting the magnitudes of the edge weights since edges attributed to small penalized partial correlations may actually correspond to substantially larger partial correlations in the unpenalized matrix. The penalized partial correlations are still useful, though, because they provide a useful method to compare the relative strengths edges within a graph or graph series. Though $R(E) = R'(E)$, the unweighted models are not affected by this bias since **A** does not incorporate magnitude of the partial correlations. Hence, R may be preferable to R' for some applications.

One important limitation of Markov networks is that they cannot specifically identify the higher order dependence structure of joint distributions complete induced subgraphs [31]. Without additional information about the population or the random variables comprising by $R(V)$, the researcher must rely on induction to determine which covariance structure is more likely to be correct. For example, a complete induced subgraph of a Markov network with $k = 4$ could be the result of six separate pair-wise correlations, a combination of one 3-way and 3 pair-wise correlations, or a single 4-way correlation. The frequent dense subgraph R_{BLA} is a case in the current study where a higher order dependence structure may exist but cannot be specified. Due to the strength of the edges, consistency of edges over time, and similarity of the

drugs in R_{BLA}, it would be reasonable to attribute the observed structure to a single 5-way correlation element, but combinations of multiple lower order interactions are possible.

The R-nets are unable to distinguish the biochemical or genetic mechanisms responsible for the observed phenotypic relationships, and individual edges may represent multiple categories of collateral resistance. However, some inferences may be reached via induction based our biological knowledge of the resistance mechanisms. Edges between drug resistances in the same class could be caused by cross-resistance mechanisms, or pleiotropic mutations affecting the common site of multiple drugs. For example, it is likely that β-lactamase enzymes, variations in penicillin-binding proteins or both are responsible for the dense subgraph of AMP, AMC, AXO, FOX, and TIO[51, 52]. Edges between drug resistances that belong to different classes could be caused by either co-resistance or pleiotropic resistance separately. Multiple different mechanisms may be present active in a single R-net, and multiple types of collateral resistance may contribute to a particular edge. The AZI-CHL edge was observed both years where AZI ∈ $R_y(V)$, and both drugs target the 50S ribosomal subunit. This relationship could be attributed to a pleiotropic mutation of the 50S subunit affecting the action of both AZI and CHL, co-resistance from genes residing on a plasmid or in the same genetic cassette, or both. Genotypic data could be used to elucidate the mechanisms underlying the observed phenotypic correlations, and the structure of R may provide insight into the genetic mechanisms leading the MDR bacteria. An important step in the validation of the current method will be to demonstrate the structure of *R* generated by a bacterial population conforms to the combinations of resistance genes present in the sampled bacteria.

The selection of ρ in the current method is of particular importance due to its influence on the generated networks. Here, $\rho = 0.25$ was selected because this value produced biologically coherent graphs in an informative range of densities around 20%. Larger or smaller values of ρ, such as the scenario presented in S3, may be appropriate for other applications. The subjectivity of the selection is mitigated since Θ is uniformly penalized, so any edges that are trivial at a given value of ρ will also be trivial at all higher values. Future work will explore additional methods for selecting ρ to improve standardization across studies.

The graphical nature of this method lends itself to simple visualization, allowing for complex relationships to be communicated clearly and additionally provides a framework for further analysis where the presence and magnitude of the partial correlations provide an outcome for evaluation via other statistical methods. Continuous covariates, including the number of gene copies present in an isolate, could be included as nodes in an R-net in addition to the drug resistances. Edges between a covariate node and resistance nodes could identify relationships similar to collateral resistance. Covariates of high degree may identify avenues for the indirect induction of AMR similar to collateral resistance, albeit indirectly. Graphical parameters for these covariate nodes such as centrality could provide additional insight into the evolution of highly resistant strains. A separate strategy for comparing R-nets is to test for differences in Ω generated by bacterial subpopulations. A number of methods for comparison of covariance matrices have been proposed [53] and may present a feasible proxy method to compare *R'* structure since *R'* is defined by Ω. Samples can be stratified using isolate date, e.g., isolate source, time interval of collection, genotype, etc., *R'* estimated for each stratum, and Ω compared across the strata. Unlike the strategies for comparison of covariates described earlier, which included covariates as nodes *within* the network, Ω comparisons estimate separate R-nets based on covariates *external* to the network and therefore must identify covariates of interest *a priori*. Additional work is needed to determine how to best evaluate covariates in R-nets, and it will be important to validate the phenotypic findings against genetic data to improve the interpretability of results.

We chose evaluate joint distributions of log-transformed MIC values, but it is common practice to dichotomize MIC values into susceptible or resistant categories based on breakpoints when analyzing resistance data [14, 49, 54]. The transformed AMR data contain more information than dichotomized results, and therefore analyses of the continuous data is more powerful than similar analyses of dichotomized data [55, 56]. The dichotomization of MIC values is also dependent on the selected breakpoints, which are based on clinically-relevant drug concentrations, and not based on the distribution of MICs within bacterial populations [57]. Breakpoints may also vary over time and geographic region [57–61]. We believe the networks developed from continuous resistance data more accurately represent the resistance relationships than networks based on more traditional dichotomized results.

Previous studies have used directed graphical models to describe the structure joint distributions of resistance in *E. coli* [22, 49, 62]. Directional edges represent types of causal relationships in the system they represent. In some cases, it is appropriate to assign directionality to AMR relationships, e.g., collateral resistance generated by *in vitro* selection for resistance to individual drugs [22] or the association between genes and phenotypical resistances [62]. In many other situations, including observational or surveillance studies of AMR where antimicrobial use is unknown, causality of an AMR relationship, if any exists, must be assumed [49]. In contrast, Markov networks are undirected and have no implied causality, which avoids the risk of incorrect assumptions regarding cause and effect.

While effective for describing and visualizing AMR relationships, the R-nets cannot provide information about univariate changes to the MICs over time. Hence, this method is intended complement current surveillance methodologies, not replace them. It was noted that the majority of MICs test ranges included at most 1 dilution above the CLSI suggested breakpoint for resistance, and some MIC test ranges had as few as two dilutions. Increasing the number of dilutions tested for MICs would capture more information and increase the accuracy of the R-nets.

Conclusions

The graphical models presented provide a novel method of mapping resistance relationships in observed in AMR surveillance data. The R-nets present a powerful and useful tool can provide insight into the evolution of MDR bacterial strains and allow simple visualization of complex AMR data. Future work is needed to validate the R-nets against genotypic results to confirm the observed phenotypic resistances accurately represent the underlying biochemical mechanisms of AMR. The application of the method to pathogenic species of bacteria, including MRSA, *Salmonella* spp. and *Campylobacter* spp., is planned and may provide insight into antimicrobials driving the evolution and emergence MDR bacteria.

Supporting Information

S1 Text. Data source and sample sizes. Summary of *E. coli* isolates and resistance testing methods for NARMS data used to demonstrate the R-net method. Includes table S1-1 and S2-2.
(DOCX)

S2 Text. Exploration of results under various L_1 penalties. An expanded evaluation of how the L_1 penalty effects the structure of R-nets. Includes tables S2-1 and S2-2.
(DOCX)

S3 Text. Alternate R-net Results. A review of the structure of the *E. coli* R-nets conditioned on $L_1 = 0.10$. This penalty was identified using the objective methods described in S2, but was

rejected in favor of a higher penalty due to difficulties in interpreting the dense networks produced here. Includes tables S3-1 and S3-2.
(DOCX)

S1 Fig. Sample size for *E. coli* by NARMS agency and year. A stacked histogram representing the 14,418 *E. coli* isolates provided to the NARMS study between 2004 and 2012 by the United States Department of Agriculture Agricultural Research Service (USDA-ARS, red portions) and Food and Drug Administration (FDA, blue portions) from chicken carcass rinsates and commercially-packed chicken breast products, respectively.
(TIF)

S2 Fig. Comparison of proportion of within- and between-class edges under multiple L_1 penalties. Edges joining resistances to drugs of a similar class ('within-class' edges) were substantially more robust to low regularization penalties ($\rho < 0.30$), while edges joining resistances of different antimicrobial classes were disproportionately removed by the lower penalties. The difference in behavior between these two types of edges led to differences in unweighted modularity based on the selected penalty. The drug resistances were grouped into one of the following classes based on drug structure: aminoglycosides, β-lactams, fluoroquinolones (including NAL), sulfonamides, tetracyclines, and macrolides (see Table 1). Network structures were estimated from MIC data for 16 drugs from 14,418 E. coli isolates collected by the FDA and USDA during 2004–12. The vertical line at $\rho = 0.25$ indicates the penalty used to generate R and R' in the presented study.
(TIF)

S3 Fig. Markov networks of AMR surveillance data under an alternate regularization penalty. Nine weighted Markov networks representing AMR data consisting of 14,418 total isolates of E. coli collected from chicken carcass rinsates and commercial chicken breasts from each year from 2004 to 2012. Networks were generated via the graphical least absolute shrinkage and selection operator with $\rho = 0.10$. In general, these graphs were considered too dense to be informative and included unstable cycles in several instances. The edges were decorated with line weights and styles to indicate the relative $|\omega_{ij}|$ magnitude and the sign of the partial correlation. Vertex size indicates the percent of isolates with an MIC meeting or exceeding the published breakpoint for respective drug (See Table 2). Vertex colors indicate the Class of drug associated resistance as follows: β-lactams–light green, fluoroquinolones–blue, aminoglycosides–purple, sulfonamides–yellow; chloramphenicol–dark green; tetracycline–red; macrolide–pink.
(TIF)

S4 Fig. Histogram of edge frequency in Markov networks of AMR surveillance data under an alternate regularization penalty. Sixty-one unique edges were identified across 9 Markov networks estimated using AMR data from 14,418 total isolates of E. coli collected from chicken carcass rinsates (USDA-ARS) and commercial chicken breasts (FDA) between 2004 and 2012. Networks were generated via the graphical least absolute shrinkage and selection operator (LASSO) with $\rho = 0.10$. A similar bimodal distribution to that seen under $\rho = 0.25$ was noted, but with a higher peak at the lower end of the plot representing infrequently present edges. Edges are categorized as either within-class (blue) or between-class (red) edges.
(TIF)

S5 Fig. Heatmap of non-zero partial correlations between drug resistances over time under an alternate regularization penalty (A) Edge weights for 17 edges between resistances of drugs of the same class and (B) 44 edges between resistances to drugs of different classes in

Markov networks of AMR data. Red and green coloring represent the magnitude of the positive and negative partial correlations, respectively, defining the edge. Grey areas represent years when one of the antibiotic drugs was not included in the panel (AMI after 2010 and AZI prior to 2011) and the edge could not be observed. The AMI-KAN edge was the only within-class edge found under $\rho = 0.10$ that did not appear under $\rho = 0.25$. The lower penalty found many more unique between-class edges (44 vs 17), but many of these were transient. Edges that appeared transiently when $\rho = 0.25$ were more consistently present under $\rho = 0.10$.
(TIF)

S6 Fig. Modularity in Markov networks of AMR surveillance data over time. The unweighted modularity (Q, blue line) and weighted modularity (Q', red line) based on the class of drug associated with the resistance (See Table 1) for Markov networks of AMR data over time. Networks were generated via the graphical least absolute shrinkage and selection operator with $\rho = 0.10$. A statistically negative trend over time was noted in both Q (Spearman's rho = -0.86, p < 0.005) and Q' (Spearman's rho = -0.96, p < 0.005).
(TIFF)

Acknowledgments

Author Contributions

Conceived and designed the experiments: WJL YTG CL.

Performed the experiments: WJL.

Analyzed the data: WJL.

Contributed reagents/materials/analysis tools: WJL KAZ JGB.

Wrote the paper: WJL.

References

1. Hawkey PM, Jones AM. The changing epidemiology of resistance. Journal of Antimicrobial Chemotherapy. 2009; 64(suppl 1):i3–i10.

2. Liu YY, Wang Y, Walsh TR, Yi LX, Zhang R, Spencer J, et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. Lancet Infect Dis. 2016; 16(2):161–8. doi: 10.1016/S1473-3099(15)00424-7 PMID: 26603172

3. Paterson DL, Harris PN. Colistin resistance: a major breach in our last line of defence. Lancet Infect Dis. 2016; 16(2):132–3. doi: 10.1016/S1473-3099(15)00463-6 PMID: 26603171

4. McGann P, Snesrud E, Maybank R, Corey B, Ong AC, Clifford R, et al. Escherichia coli Harboring mcr-1 and blaCTX-M on a Novel IncF Plasmid: First Report of mcr-1 in the United States. Antimicrob Agents Chemother. 2016; 60(7):4420–1. PubMed Central PMCID: PMCPMC4914657. doi: 10.1128/AAC.01103-16 PMID: 27230792

5. Nikaido Y, Nakashima T. Effects of environmental novelty on fear-related behavior and stress responses of rats to emotionally relevant odors. Behav Brain Res. 2009; 199(2):241–6. doi: 10.1016/j.bbr.2008.11.043 PMID: 19103229

6. Baquero F. From pieces to patterns: evolutionary engineering in bacterial pathogens. Nat Rev Microbiol. 2004; 2(6):510–8. doi: 10.1038/nrmicro909 PMID: 15152207

7. Blumberg HM, Rimland D, Carroll DJ, Terry P, Wachsmuth IK. Rapid development of ciprofloxacin resistance in methicillin-susceptible and -resistant Staphylococcus aureus. J Infect Dis. 1991; 163 (6):1279–85. PMID: 2037793

8. Canton R, Ruiz-Garbajosa P. Co-resistance: an opportunity for the bacteria and resistance genes. Curr Opin Pharmacol. 2011; 11(5):477–85. doi: 10.1016/j.coph.2011.07.007 PMID: 21840259

9. Cullen ME, Wyke AW, Kuroda R, Fisher LM. Cloning and characterization of a DNA gyrase A gene from Escherichia coli that confers clinical resistance to 4-quinolones. Antimicrob Agents Chemother. 1989; 33(6):886–94. PubMed Central PMCID: PMCPMC284251. PMID: 2548439

10. Ariza AC, Linares P, de Castro MD, Valcarcel M. Unsegmented flow approach for on-line monitoring of pH, conductivity, dissolved oxygen and determination of nitrite and ammonia in aquaculture. J Automat Chem. 1994; 16(2):59–62. PubMed Central PMCID: PMCPMC2562827. doi: 10.1155/S1463924694000052 PMID: 18924685

11. Cohen SP, McMurry LM, Hooper DC, Wolfson JS, Levy SB. Cross-resistance to fluoroquinolones in multiple-antibiotic-resistant (Mar) Escherichia coli selected by tetracycline or chloramphenicol: decreased drug accumulation associated with membrane changes in addition to OmpF reduction. Antimicrob Agents Chemother. 1989; 33(8):1318–25. PubMed Central PMCID: PMCPMC172647. PMID: 2679373

12. CDC. Antibiotic Resistance Threats in the United States Washington; D.C.2013. Available from: http://www.cdc.gov/drugresistance/threat-report-2013/index.html.

13. Antimicrobial Resistance Action Plan. In: Agriculture USDo, editor. Washington; DC: United States Department of Agriculture; 2014. p. pp. 28.

14. FDA. NARMS Integrated Report: 2012–2013. In: Administration FaD, editor. 2015.

15. Ashiru-Oredope D, Utilization SHobotESPfA, Group RO. Antimicrobial stewardship: English Surveillance Programme for Antimicrobial Utilization and Resistance (ESPAUR). Journal of Antimicrobial Chemotherapy. 2013; 68(11):2421–3. doi: 10.1093/jac/dkt363 PMID: 24027247

16. Updated Public Health Microbiology Strategy & Work Plan 2012–2016. In: Control ECfDPa, editor. 2012. p. pp 21.

17. ECDC. Long-term surveillance strategy 2014–2020. Stockholm; Sweden: European Centre for Disease Prevention and Control; 2013. 24 p.

18. 2009 EU—U.S. Summit Declaration2009. Available from: http://www.cdc.gov/drugresistance/pdf/nov-3-2009-summit-declaration_en0_508.pdf.

19. CDC. Transatlantic Taskforce on Antimicrobial Resistance: Progress Report. TATFAR; 2014.

20. EFSA. EUSR on AMR in zoonotic and indicator bacteria from humans, animals and food 2014. European Food Safety Authority Journal. 2016; 14(2):pp. 207.

21. Baym M, Stone LK, Kishony R. Multidrug evolutionary strategies to reverse antibiotic resistance. Science. 2016; 351(6268):aad3292. doi: 10.1126/science.aad3292 PMID: 26722002

22. Imamovic L, Sommer MO. Use of collateral sensitivity networks to design drug cycling protocols that avoid resistance development. Sci Transl Med. 2013; 5(204):204ra132. doi: 10.1126/scitranslmed.3006609 PMID: 24068739

23. Wagner BA, Salman MD, Dargatz DA, Morley PS, Wittum TE, Keefe TJ. Factor analysis of minimum-inhibitory concentrations for Escherichia coli isolated from feedlot cattle to model relationships among antimicrobial-resistance outcomes. Prev Vet Med. 2003; 57(3):127–39. PMID: 12581596

24. FDA. NARMS Now: Integrated Data [cited 2016 Mar 31]. Available from: http://www.fda.gov/AnimalVeterinary/SafetyHealth/AntimicrobialResistance/NationalAntimicrobialResistanceMonitoringSystem/ucm458213.htm.

25. Team RC. R: A Language and Environment fo Statistical Computing. 3.2.3 ed. Vienna, Austria: R Foundation for Statistical Computing; 2016.

26. Spearman C. The Proof and Measurement of Association between Two Things. The American Journal of Psychology. 1904; 15(1):72–101.

27. Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics. 2008; 9(3):432–41. PubMed Central PMCID: PMCPMC3019769. doi: 10.1093/biostatistics/kxm045 PMID: 18079126

28. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, Complex Systems. 2006; 1695(5):1–9.

29. Newman MEJ. Networks: an introduction. Oxford; New York: Oxford University Press; 2010. p. xi, 772 p.

30. Kindermann R, Snell JL, American Mathematical Society. Markov random fields and their applications. Providence, R.I.: American Mathematical Society; 1980. ix, 142 p. p.

31. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. Springer series in statistics,. 2nd ed. New York, NY: Springer; 2009. p. xxii, 745 p.

32. Peng J, Wang P, Zhou N, Zhu J. Partial Correlation Estimation by Joint Sparse Regression Models. Journal of the American Statistical Association. 2009; 104(486):735–46. doi: 10.1198/jasa.2009.0126 PMID: 19881892

33. Liu H, Han F, Yuan M, Lafferty J, Wasserman L. High-dimensional semiparametric Gaussian copula graphical models. 2012:2293–326.

34. Cartwright D, Harary F. Structural balance: a generalization of Heider's theory. Psychol Rev. 1956; 63 (5):277–93. Epub 1956/09/01. PMID: 13359597

35. Cattell RB. The Scree Test For The Number Of Factors. Multivariate Behav Res. 1966; 1(2):245–76. doi: 10.1207/s15327906mbr0102_10 PMID: 26828106

36. D'Agostino RB, Russell HK. Scree Test. Encyclopedia of Biostatistics: John Wiley & Sons, Ltd; 2005.

37. Newman MEJ. Networks: an introduction. Oxford; New York: Oxford University Press; 2010.

38. Newman ME, Girvan M. Finding and evaluating community structure in networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2004; 69(2 Pt 2):026113.

39. Gomez S, Jensen P, Arenas A. Analysis of community structure in networks of correlated data. Phys Rev E Stat Nonlin Soft Matter Phys. 2009; 80(1 Pt 2):016114. Epub 2009/08/08.

40. Kaiser HF. The Application of Electronic Computers to Factor Analysis. Educational and Psychological Measurement. 1960; 20(1):141–51.

41. Clarkson DB, Jennrich RI. Quartic rotation criteria and algorithms. Psychometrika. 1988; 53(2):251–9.

42. Guadagnoli E, Velicer WF. Relation of sample size to the stability of component patterns. Psychol Bull. 1988; 103(2):265–75. PMID: 3363047

43. Meinshausen N, Buhlmann P. High-dimensional graphs and variable selection with the Lasso. 2006:1436–62.

44. Yamagishi J, Yoshida H, Yamayoshi M, Nakamura S. Nalidixic acid-resistant mutations of the gyrB gene of Escherichia coli. Molecular & general genetics: MGG. 1986; 204(3):367–73. Epub 1986/09/01.

45. Breines DM, Ouabdesselam S, Ng EY, Tankovic J, Shah S, Soussy CJ, et al. Quinolone resistance locus nfxD of Escherichia coli is a mutant allele of the parE gene encoding a subunit of topoisomerase IV. Antimicrob Agents Chemother. 1997; 41(1):175–9. PubMed Central PMCID: PMCPMC163680. PMID: 8980775

46. Khodursky AB, Zechiedrich EL, Cozzarelli NR. Topoisomerase IV is a target of quinolones in Escherichia coli. Proc Natl Acad Sci U S A. 1995; 92(25):11801–5. PubMed Central PMCID: PMCPMC40490. PMID: 8524852

47. Sorlozano A, Gutierrez J, Jimenez A, de Dios Luna J, Martinez JL. Contribution of a new mutation in parE to quinolone resistance in extended-spectrum-beta-lactamase-producing Escherichia coli isolates. J Clin Microbiol. 2007; 45(8):2740–2. Epub 2007/06/15. PubMed Central PMCID: PMCPMC1951223. doi: 10.1128/JCM.01093-07 PMID: 17567787

48. George AM, Levy SB. Amplifiable resistance to tetracycline, chloramphenicol, and other antibiotics in Escherichia coli: involvement of a non-plasmid-determined efflux of tetracycline. J Bacteriol. 1983; 155 (2):531–40. PubMed Central PMCID: PMCPMC217720. PMID: 6348022

49. Ludwig A, Berthiaume P, Boerlin P, Gow S, Leger D, Lewis FI. Identifying associations in Escherichia coli antimicrobial resistance patterns using additive Bayesian networks. Prev Vet Med. 2013; 110 (1):64–75. doi: 10.1016/j.prevetmed.2013.02.005 PMID: 23484804

50. Hastie T, Tibshirani R, Wainwright M. Statistical learning with sparsity: the lasso and generalizations. Monographs on statistics and applied probability. Boca Raton: CRC Press, Taylor & Francis Group; 2015. p. xv, 351 pages.

51. Bush K, Jacoby GA, Medeiros AA. A functional classification scheme for beta-lactamases and its correlation with molecular structure. Antimicrobial Agents and Chemotherapy. 1995; 39(6):1211–33. PMID: 7574506

52. Spratt BG. Properties of the Penicillin-Binding Proteins of Escherichia coli K12. European Journal of Biochemistry. 1977; 72(2):341–52. PMID: 319999

53. Steppan SJ, Phillips PC, Houle D. Comparative quantitative genetics: evolution of the G matrix. Trends in Ecology & Evolution. 2002; 17(7):320–7.

54. FDA. National Antimicrobial Resistance Monitoring System Executive Report: 2011. 2014.

55. Naggara O, Raymond J, Guilbert F, Roy D, Weill A, Altman DG. Analysis by categorizing or dichotomizing continuous variables is inadvisable: an example from the natural history of unruptured aneurysms. AJNR Am J Neuroradiol. 2011; 32(3):437–40. doi: 10.3174/ajnr.A2425 PMID: 21330400

56. Fedorov V, Mannino F, Zhang R. Consequences of dichotomization. Pharmaceutical Statistics. 2009; 8(1):50–61. doi: 10.1002/pst.331 PMID: 18389492

57. CLSI. Performance Standards for Antimicrobial Susceptibility Testing. Wayne, PA, USA: Clinical Laboratory and Standards Institute; 2016. Available from: http://em100.edaptivedocs.com/.

58. Hombach M, Mouttet B, Bloemberg GV. Consequences of revised CLSI and EUCAST guidelines for antibiotic susceptibility patterns of ESBL- and AmpC beta-lactamase-producing clinical Enterobacteriaceae isolates. J Antimicrob Chemother. 2013; 68(9):2092–8. doi: 10.1093/jac/dkt136 PMID: 23633681

59. Hamada Y, Sutherland CA, Nicolau DP. Impact of Revised Cefepime CLSI Breakpoints on Escherichia coli and Klebsiella pneumoniae Susceptibility and Potential Impact If Applied to Pseudomonas aeruginosa. Journal of Clinical Microbiology. 2015; 53(5):1712–4. doi: 10.1128/JCM.03652-14 PMID: 25694527

60. Dudley MN, Ambrose PG, Bhavnani SM, Craig WA, Ferraro MJ, Jones RN, et al. Background and Rationale for Revised Clinical and Laboratory Standards Institute Interpretive Criteria (Breakpoints) for Enterobacteriaceae and Pseudomonas aeruginosa: I. Cephalosporins and Aztreonam. Clinical Infectious Diseases. 2013; 56(9):1301–9. doi: 10.1093/cid/cit017 PMID: 23334813

61. EUCAST. Breakpoint tables for interpretation of MICs and zone diameters, v6.0. European Committee on Antimicrobial Susceptibility Testing, 2015.

62. Hidano A, Yamamoto T, Hayama Y, Muroga N, Kobayashi S, Nishida T, et al. Unraveling antimicrobial resistance genes and phenotype patterns among Enterococcus faecalis isolated from retail chicken products in Japan. PLoS One. 2015; 10(3):e0121189. PubMed Central PMCID: PMCPMC4363150. doi: 10.1371/journal.pone.0121189 PMID: 25781022

63. USDA-ARS. National Antimicrobial Resistance Monitoring System Animal Isolates [cited 2016 Mar 31]. Available from: http://www.ars.usda.gov/Main/docs.htm?docid=6750&page=3.