An accurate and interpretable model for antimicrobial resistance in pathogenic *Escherichia coli* from livestock and companion animal species

Henri C. Chung[1], Christine L. Foxx[2], Jessica A. Hicks[3], Tod P. Stuber[3], Iddo Friedberg[1], Karin S. Dorman[4,5], Beth Harris[6*]

1 Department of Veterinary Microbiology and Preventive Medicine, Iowa State University, Ames IA, USA 50011
2 Research Participation Program, Oak Ridge Institute for Science and Education, Oak Ridge, TN, USA 37830
3 National Veterinary Services Laboratories, Animal and Plant Health Inspection Service, U.S. Department of Agriculture, Ames, IA, USA 50010
4 Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA, USA 50011
5 Department of Statistics, Iowa State University, Ames, IA, USA 50011
6 National Animal Health Laboratory Network, National Veterinary Services Laboratories, Animal and Plant Health Inspection Service, U.S. Department of Agriculture, Ames, IA, USA 50010

* beth.n.harris@usda.gov

Abstract

Understanding the microbial genomic contributors to antimicrobial resistance (AMR) is essential for early detection of emerging AMR infections, a pressing global health threat in human and veterinary medicine. Here we used whole genome sequencing and antibiotic susceptibility test data from 980 disease causing *Escherichia coli* isolated from companion and farm animals to model AMR genotypes and phenotypes for 24 antibiotics. We determined the strength of genotype-to-phenotype relationships for 197 AMR genes with elastic net logistic regression. Model predictors were designed to evaluate different potential modes of AMR genotype translation into resistance phenotypes. Our results show a model that considers the presence of individual AMR genes and total number of AMR genes present from a set of genes known to confer resistance was able to accurately predict isolate resistance on average (mean F_1 score = 98.0%, SD = 2.3%, mean accuracy = 98.2%, SD = 2.7%). However, fitted models sometimes varied for antibiotics in the same class and for the same antibiotic across animal hosts, suggesting heterogeneity in the genetic determinants of AMR resistance. We conclude that an interpretable AMR prediction model can be used to accurately predict resistance phenotypes across multiple host species and reveal testable hypotheses about how the mechanism of resistance may vary across antibiotics within the same class and across animal hosts for the same antibiotic.

Keywords

antibiotics, antimicrobial resistance, *Escherichia coli*, genotype, phenotype, surveillance, whole-genome sequencing, genotype-phenotype association

Introduction

Antimicrobial resistance (AMR) is one of the foremost concerns in human and animal health and well-being. Multiple organizations including the World Health Organization, the European Commission, and major U.S. agencies such as the U.S. Centers for Disease Control and Prevention (CDC), the U.S. Food and Drug Administration (FDA), and the U.S. Department of Agriculture (USDA), have recognized the global threat of AMR infections [1–3]. AMR infections are estimated to cost more than US$20–35 billion per year in increased clinical treatment costs, and to contribute an additional 35,000 deaths per year in the U.S. alone [3, 4]. In response to this threat, the U.S. government has issued in 2015 the first National Action Plan for Combating Antibiotic-Resistant Bacteria (CARB NAP), directing all U.S. Federal agencies to collaboratively develop strategies against AMR infection using an integrated One Health approach [5]. The One Health approach recognizes the interconnection between humans, plants, animals, and the environment in global health, and emphasizes using information across sectors to understand and combat AMR [6]. CARB NAP includes the incorporation of multiple data streams such as the National Antimicrobial Resistance Monitoring System (NARMS), recommendation strategies for the judicious use of antibiotics in veterinary medicine, and restrictions on the use of certain glycopeptides, fluoroquinolones, and cephalosporins in both human and animal healthcare [7, 8].

As part of the USDA response to the CARB NAP, the Animal and Plant Health Inspection Service's National Animal Health Laboratory Network (APHIS-NAHLN) established a collaborative AMR pilot project in 2018 with American Association of Veterinary Laboratory Diagnosticians (AAVLD) member laboratories [9]. A working group consisting of representatives from the AAVLD member laboratories, Clinical Laboratory Standards Institute (CLSI), the FDA Center for Veterinary Medicine Veterinary Laboratory Investigation and Response Network (VetLIRN), and the USDA

Center for Epidemiology and Animal Health (CEAH) was convened to recommend methods and standards that would meet the pilot project's specific aims. These project aims included: (1) development of a sampling stream to monitor AMR in animal pathogens routinely isolated in veterinary diagnostic laboratories; (2) development of standardized methods for sharing and disseminating this information within the veterinary community, and (3) making data publicly accessible to inform future policies designed to mitigate AMR in animals [9]. Here we capitalized on the unique and extensive data from this pilot project to assess the state of knowledge and further elucidate the genetic mechanisms of AMR in this diverse set of agricultural and companion animals.

Genetic models of AMR are an important area of veterinary research. The decreased cost and increased availability of genome sequencing technologies has lead to a vast increase in the known genetic determinants of AMR, and the subsequent creation of over a dozen reference genetic databases and gene detection tools to assist AMR identification and research [10]. This information is then used to support diagnostic and surveillance observations made by traditional culture-based methods, and improve identification in cases where bacteria cannot be cultivated [11, 12]. Recently, researchers seeking to capitalize on this rapid growth of information have turned to machine learning (ML) models to further enhance AMR surveillance [13]. ML models have been successfully used to predict AMR resistance in pathogens such as non-typhoidal *Salmonella*, *Mycobacterium tuberculosis*, and *E. coli* [14–21]. These models use sequencing information to make accurate predictions of resistance phenotype, alleviating the need for antimicrobial susceptibility testing and potentially identifying new AMR genetic determinants in the process [22].

Due to this increasing interest, there is a large number of studies on genotype-to-phenotype relationships and ML-based prediction of AMR. However, there is little previous research on how these relationships vary depending on the specific antibiotic, animal host, or other factors. While several studies have found 95% to 100% correlation between AMR genotype and phenotype in multiple bacterial pathogens, these studies also find appreciable differences in the rates of resistance to fluoroquinolones, tetracyclines, aminoglycosides and β-lactams [23–25]. The differences in rates of resistance may be attributed to underlying differences in the resistance mechanisms of AMR gene products or may simply be due to some of them used exclusively in veterinary or animal husbandry practice [26, 27]. Additionally, bacterial isolates express widely variable rates of resistance depending on the host animal species [28, 29]. However, animal host information has rarely been considered in AMR models, and rarely do studies consider these relationships in more than one host animal at a time [30, 31]. Some differences in the micro-environments found in different animal hosts have been found to be a significant source of genetic variation [32–34]. This variation suggests that AMR genotype-phenotype relationships could be modeled in a more complex fashion. A meta-analysis examining the performance and reliability of ML models for AMR prediction in *Neisseria gonorrhoeae*, *Klebsiella pneumoniae* and *Acinetobacter baumannii* showed that model performance can vary significantly depending on the antibiotic, data set, resistance metric, and bacterial species [35]. Although ML methods in AMR prediction have been shown to be accurate, there is a need to better understand the variation in predictions and incorporate relevant biological and clinical knowledge into model design and evaluation before adoption in large-scale use.

Here we used sequenced isolates of *E. coli*, a zoonotic and cross-species pathogen, to indicate emerging AMR [36, 37]. *E. coli* may serve as a reservoir for transferring AMR genes to other animal and human populations through plasmids, transposons, and other mobile genetic elements [38]. Therefore, by examining the AMR profiles of *E. coli*

isolates, we aim to develop an understanding of the different genetic associations that lead to the expression of AMR phenotypes in domesticated animals of interest to the clinical veterinary diagnostic community and public health practitioners.

We used genotype data from 980 *E. coli* isolates recovered from seven animal species to determine the utility of AMR genotype data for predicting phenotypic resistance in a clinically relevant context. Isolates came from various species of agricultural and companion animals as part of the APHIS-NAHLN AMR pilot project from 2018-2022. Our goal was to construct an interpretable, biologically informed model that would consider AMR gene content and host animal species to predict resistance to specific antibiotics. To do so, we fit and compared multiple elastic net models to characterize the relationship between AMR genes and predict resistance phenotype. We considered predictors that test different ways in which AMR genes could confer resistance. The proposed associations based on features that include the presence or absence of at least one gene within a group, the number of genes within each group, and the host animal effects. We then evaluated each model's performance by its ability to predict AMR phenotypes. Our results show that while the different models have high predictive performance, there are significant differences in feature importance in predicting resistance to specific antibiotics in specific hosts. These differences suggest that future AMR modeling approaches could be made more accurate by incorporating information for each antibiotic and animal host. We discuss these differences in terms of the prevalence of resistance between different host animals, model complexity, gene association (binary presence/absence or summative count), and unique predictors. We also provide an estimate of our best fit model's power to identify predictors of AMR phenotype in this study.

Materials and Methods

The outline of this work is shown in Fig 1.

Laboratory enrollment

Laboratories participating in the AMR pilot project were enrolled annually by APHIS-NAHLN. To maximize the representation of data at a national level, we considered factors such as animal population, representation, and geography in the enrollment process. Laboratories were enrolled and submitted the sequence data represented in this analysis. At the time of writing, the 27 participating laboratories were located in the following states: Alabama, California, Colorado, Florida, Georgia, Indiana, Iowa, Kansas, Kentucky, Louisiana, Michigan, Minnesota, Mississippi, Missouri, Nebraska, New York, North Carolina, North Dakota, Ohio, Pennsylvania, South Dakota, Texas, Washington, and Wisconsin.

Isolate selection

We received data from 981 total *E. coli* samples from swine (102), cattle (198), chickens (121), turkeys (42), ducks (1), horses (151), dogs (195), and cats (173). Participating laboratories selected isolates for inclusion in the pilot study based on association with clinical disease, with one sample per distinct animal source (farm, herd, or owner within a single year). Each laboratory assigned a unique identifier to an isolate prior to submitting to APHIS-NAHLN, to eliminate all identifiable information from the data. Isolates were identified at the genus and species level by each laboratory, typically by using MALDI-TOF methods standard for veterinary microbiology laboratories [39]. Data submitted for each isolate included the minimum inhibitory concentration (MIC)

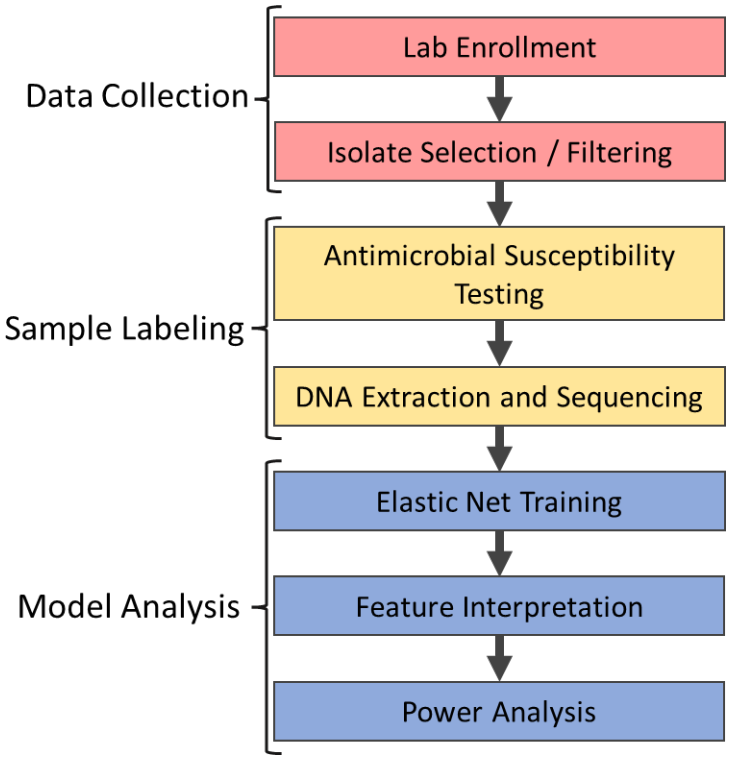

Fig 1. Project Workflow. We separate the project workflow into three distinct phases. In Data Collection (red); we enrolled participating USDA veterinary laboratories and collected *E. coli* isolates from animals associated with a clinical disease. We removed samples with improper or insufficient metadata. We then labeled samples (yellow) with their corresponding AMR phenotype and genotype using AST and WGS (see Methods). In the final model analysis (blue), we fit the phenotype and genotype information to multiple elastic net models, interpreted the importance of model features, and conducted a power analysis to assess the detectable effect sizes given the study sample sizes.

values for all antibiotics tested, date of isolation, host animal species, specimen/tissue source, and final clinical diagnosis where available. After cleaning the data entries for typographical errors, we removed the duck sample from subsequent analyses as it represented a single animal from that species. Tissue samples were originally collected for other clinical purposes.

Antimicrobial susceptibility testing

Participating laboratories conducted all *E. coli* antimicrobial susceptibility testing (AST) using the following commercially available Sensititre™plates (Thermofisher Scientific, USA) according to the manufacturer's instructions: BOPO6F or BOPO7F (swine and cattle), EQUIN1F or EQUIN2F (horses), AVIAN1F (chickens and turkeys), and COMPGN1F (dogs and cats). The six plate layouts previously listed include a total of 50 antibiotics from 14 different antibiotic classes, with each plate type containing between 18-24 different antibiotics.

We used current clinical breakpoint guidelines from the Clinical and Laboratory Standards Institute (CLSI) established based on a combination of factors including the host animal species, sample source and type (*e.g.*, urinary tract or skin/soft tissue infection), and bacterial isolate. Seventeen antibiotics have interpretive clinical breakpoints established for *E. coli* in various animal species at the time of writing in the *Vet01S* [40]. These were: amikacin (dogs, horses), amoxicillin-clavulanic acid (cats, dogs), ampicillin (cattle, cats, dogs, horses), cefazolin (dogs, horses), cefovecin (cats,

dogs), cefpodoxime (dogs), ceftazidime (dogs), ceftiofur (cattle, swine), cephalexin (dogs), doxycycline (horses), enrofloxacin (cats, dogs, horses, poultry), gentamicin (dogs, horses), marbofloxacin (cats, dogs), minocycline (horses), orbifloxacin (cats, dogs), piperacillin-tazobactam (dogs), and pradofloxacin (cats, dogs). Based on these breakpoints, we classified samples as expressing sensitive, intermediate, or resistant antimicrobial phenotypes (Table S1). To simplify model prediction, intermediate phenotypes were labeled as antibiotic resistant.

For antimicrobials without CLSI breakpoints, samples were classified into wildtype (*wt*) and non-*wt* groups using epidemiological cut-off values (ECOFF) from the European Committee on Antimicrobial Susceptibility (EUCAST) [41]. Isolates with MIC values at or below the ECOFF value are considered to be members of the wildtype population for that given bacterial species and antimicrobial. ECOFF values are calculated by fitting a cumulative log-normal distribution using non-linear least squares regression to MIC data curated by the EUCAST database [42]. ECOFF values were collected from the EUCAST MIC distribution website (http://www.eucast.org) in February 2023. Unlike CLSI breakpoints, ECOFF values are non-host animal specific. If an isolate's phenotype was ambiguous, such as MIC values reported as "less than or equal" to a value exceeding its CLSI breakpoint or ECOFF value, the sample phenotype was labeled "Non-Interpretable". Isolate and antimicrobial combinations without a CLSI breakpoint or ECOFF value, or with a Non-Interpretable phenotype, were removed from further analysis (Table S2).

DNA extraction and whole-genome sequencing

Isolates were either sequenced directly by participating laboratories or submitted to the National Veterinary Services Laboratories (NVSL) for whole-genome sequencing (WGS). Briefly, extracted DNA was used to prepare indexed genomic libraries using the Nextera XT® DNA Library Prep Kit (Illumina). Multiplexed libraries were sequenced using 250×2 paired-end read chemistry on the Illumina MiSeq® platform at an average sequencing depth of 91.2 ± 32.3-fold.

Isolates were verified as *E. coli* using Kraken against a database consisting of RefSeq complete genomes database release ver. 2.09 [43], UniVec-core, and host genomes commonly encountered by the NVSL, with 85% reference genome coverage or greater after Bayesian Reestimation of Abundance, using default parameters [44, 45]. Following this validation step, *de novo* assembly of bacterial genomes was performed using SPAdes v3.14.0 [46] and resultant scaffolds were processed using the AMRFinder v1.0.1 [47] and ABRicate v3.2.3 [48] tool kits against the National Center for Biotechnology Information (NCBI), AMRFinder, and ResFinder [49] databases. Resultant gene hits from these tool kits, with minimum inclusion requirements of 70% amino-acid identity and 95% reference gene coverage, were then collated with the metadata as putative AMR genes found in each *E. coli* isolate. Plasmid genes were not used due to insufficient information linking the presence or absence of those genes with specific antibiotic resistance. In total, 197 different AMR genes were identified across all samples.

Determining animal host differences

We used Fisher's exact test to determine if there were significant differences in the fraction of resistant or non-*wt* phenotypes between animal hosts for each antibiotic in the data set. We applied the Bonferroni-Holm (BH) correction for multiple comparisons to p-values, and a BH corrected p-value threshold of 0.05 was used to determine significance. All analyses and statistical tests were done using R 4.2.2, model training and evaluation was performed using the tidymodels family of packages, and visualizations created with the ggplot2 3.4.1 [50–52].

Elastic net regularization

We used an elastic net model to estimate the effect of relevant genotype predictors on the resistance phenotype. An elastic net is a penalized regression model which combines the penalties of Least Absolute Shrinkage and Selection Operator (LASSO) and Ridge regressions [53–55]. The elastic net utilizes two tunable parameters, a shrinkage parameter λ for the LASSO regression, and a mixing parameter α that determines the relative importance of the LASSO and Ridge penalties when they are added together. Rifge-driven shrinkage and LASSO-driven predictor selection make elastic net suitable for AMR prediction from genotype, where the number of predictors greatly exceeds the number of samples, while retaining a higher level of predictor interpretability compared with other ML methods [56]. The candidate elastic net predictors were built from relevant antibiotic genes, defined as the set of AMR genes known to cause resistance to the specified antibiotic.

We curated a custom database that captures the knowledge of gene-specific AMR based on common published literature (See Data Availability). In this database, AMR genes are associated with specific individual antibiotics or a general class of antibiotics. Throughout this analysis, we separately fit samples with CLSI and ECOFF breakpoints. In addition, to avoid fitting models with low power, we excluded antibiotics represented in fewer than 5 samples per phenotype label. For the remaining 31 antibiotic and breakpoint combinations, we only admitted genes into the relevant set $\mathcal{X}_i$ that are known to confer resistance to antibiotic i. The exception was the *Full* model, where all AMR genes, including genes associated with other antibiotics, were considered. We then used elastic net with the logistic regression (Equation (1)) to model and predict isolate antibiotic resistance based on the presence or absence of genes in the gene group.

$$\Pr(y_{ij} = 1 \mid \boldsymbol{x}_{ij}, \boldsymbol{w}_{ij}) = \frac{1}{1 + e^{\boldsymbol{x}_{ij}^t \boldsymbol{\beta}_i + \boldsymbol{w}_{ij}^t \boldsymbol{\alpha}_i + \boldsymbol{x}_{ij}^t \boldsymbol{\Gamma}_i \boldsymbol{w}_{ij}}} \tag{1}$$

In equation (1), y_{ij} is an element in the binary vector $\boldsymbol{y}_i^t = (y_{i1}, y_{i2}, \ldots, y_{im_i})$, where $y_{ij} = 1$ denotes resistance or non-*wt* response to the i-th of n antibiotics in the j-th of m_i samples. Vector $\boldsymbol{x}_{ij}^t = (x_{ij1}, x_{ij2}, \ldots, x_{ijp_i})$, where $x_{ijk} \in \{0, 1\}$ represents the presence or absence in the j-th sample of the k-th of p_i predictors which confer resistance or non-*wt* phenotype to the ith antibiotic. Vector $\boldsymbol{\beta}_i^t = (\beta_{i1}, \beta_{i2}, \ldots, \beta_{ip_i})$ denotes the set of regression coefficients for the gene effects to antibiotic i, each β_{ik} a measure of the relative degree of association between a gene k and the resistance or non-*wt* phenotype to antibiotic i. To account for potential effects of the animal host on the rate of isolate response, we include a set of predictors, $\boldsymbol{w}_{ij}^t = (w_{ij1}, w_{ij2}, \ldots, w_{ijl_i})$ corresponding to a one-hot encoding of the animal host for each isolate j tested for resistance or non-*wt* phenotype to antibiotic i and coefficients $\boldsymbol{\alpha}_i^t = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{il_i})$ for the effect of each animal on resistance or non-*wt* phenotype to antibiotic i among l_i distinct species treated with this antibiotic. For example, in a model for antibiotic i with data only from swine and cattle samples, $\boldsymbol{w}_{ij}$ is the length-two predictor for "swine" and "cattle" with 0 or 1 values to indicate if sample j came from a swine or cattle host. An interaction term, $\boldsymbol{x}_{ij}^t \boldsymbol{\Gamma}_i \boldsymbol{w}_{ij}$, represents the animal-specific differences in the genetic effects for resistant or non-*wt* phenotypes in different animal hosts, with $\boldsymbol{\Gamma}_i$ a $p_i \times l_i$ matrix of coefficients. To keep notation simple, we neglect adding an extra index for CLSI or ECOFF breakpoint, but some antibiotics participate in two fits with distinct estimates for each type of breakpoint. We also considered the possible effect location, the US state or region from which a sample was sourced, may have on phenotype. We found location and host animal labels were not independent ($\chi^2 < 2.2 \times 10^{-16}$), likely due to a bias for certain animal hosts to be raised in specific geographic regions, and we therefore only considered animal hosts as a relevant predictor.

To consider the relative predictive power of different predictors, we built models

from sets of predictors implying distinct biological mechanisms of resistance, (Table 1), by default allowing main and gene interaction effects for host species. *Main effects* are those contributed by individual predictors while *interaction effects* are those in which the effect of a predictor is dependent upon one or more other predictors. The *Gene* model considers presence/absence of any gene in the relevant gene set $\mathcal{X}_i$ known to confer resistance to the i-th antibiotic. The *Binary* model considers a single binary indicator $v_{ij} = 1$ if $\sum_{k=\in\mathcal{X}_i} x_{ijk} \geq 1$, which will be predictive if the presence of any one gene from the relevant gene set $\mathcal{X}_i$ is capable of conferring resistance. The *Count* model includes only the number $z_{ij} = \sum_{k\in\mathcal{X}_i} x_{ijk}$ of observed relevant resistance genes as a predictor, implying each additional relevant gene from $\mathcal{X}_i$ linearly increases the log odds of resistance. We also fit various combinations of these models. Lastly, we tested a *Full* model using all AMR genes identified in an isolate, ignoring biological knowledge about whether those genes are known to confer resistance to the predicted antibiotic. We define candidate predictors as the set of predictors used in each model.

Table 1. Elastic net model formulations. We use figurative equations to represent the different models, using i to index the current antibiotic in question. Covariates $\boldsymbol{w}_i$ indicate host animal, $\boldsymbol{x}_i$ indicate presence/absence of genes in the relevant set $\mathcal{X}_i$ ($\boldsymbol{x}$ for all AMR genes, not just those in the relevant set), $\boldsymbol{v}_i$ indicate presence/absence of *any* relevant gene, and $\boldsymbol{z}_i$ the count of observed relevant genes. Operator $*$ indicates interactions considered, while $+$ indicates main effects only. See text for a detailed explanation of each model.

Model	Formulation	Brief explanation
Gene	$\boldsymbol{y}_i \sim \boldsymbol{w}_i * \boldsymbol{x}_i$	Is the relevant set ($\mathcal{X}_i$) of AMR genes predictive of resistance?
Gene & Binary	$\boldsymbol{y}_i \sim \boldsymbol{w}_i * (\boldsymbol{x}_i + \boldsymbol{v}_i)$	Does knowing any gene in a given set in addition to the relevant gene set ($\mathcal{X}_i$) improve resistance prediction?
Gene & Count	$\boldsymbol{y}_i \sim \boldsymbol{w}_i * (\boldsymbol{x}_i + \boldsymbol{z}_i)$	Does knowing the total number of genes (rather than individual identity) in addition to the relevant gene set ($\mathcal{X}_i$) improve resistance prediction?
Gene & Binary & Count	$\boldsymbol{y}_i \sim \boldsymbol{w}_i * (\boldsymbol{x}_i + \boldsymbol{v}_i + \boldsymbol{z}_i)$	A combination of the two previous models.
Full	$\boldsymbol{y}_i \sim \boldsymbol{w}_i * \boldsymbol{x}$	Is the total detectable set of AMR genes, ignoring known conferred resistance, a better predictor of resistance?
Count	$\boldsymbol{y}_i \sim \boldsymbol{w}_i * \boldsymbol{z}_i$	Is the total number of genes in the relevant set ($\mathcal{X}_i$) predictive of resistance?
Binary	$\boldsymbol{y}_i \sim \boldsymbol{w}_i * \boldsymbol{v}_i$	Is the presence of any gene within the relevant set ($\mathcal{X}_i$) predictive of resistance?

Model training and feature stability

After applying the inclusion criteria, we were left with 31 subsets of data meeting our requirements: 15 using CLSI breakpoints and 16 using ECOFF values, representing 24 unique antibiotics in total. We then fit separate elastic net models to each subset of data. For each model, we further split the subset of data into 80:20 training and testing

splits, and fit the model with the training data. Within the training subset, a 10-fold cross validation scheme was used to select the value of the α and λ hyperparameters, optimizing the accuracy of predicted resistance phenotype. We evaluated the model fit by calculating the prediction accuracy of the fitted model with the testing data. When the number of candidate predictors is large, selected predictors for models can be unstable and vary between model fits [57, 58]. To identify important predictors, we used a stability selection method where we repeatedly refit the model (80:20 cross-validation and selection of α and λ) with a random 80% of candidate predictors over 1000 replicates [59, 60]. We considered important predictors to be selected predictors found in over 66% of model replicates. After training and stability selection, $\hat{\mathcal{X}}_i \subseteq \mathcal{X}_i$ denotes the set of important predictors consistently included in the model.

Model performance and power analysis

To evaluate model performance, we used a suite of evaluation metrics standard to binary classification models described in Table 2. Briefly, these different metrics discriminate the performance of the model to accurately identify samples sensitive or resistant to antibiotics, while considering the samples that are incorrectly predicted to be sensitive or resistant.

Table 2. Model Performance Metrics. Model performance was evaluated using F_1 score, Negative Predictive Value (NPV), Positive Predictive Value (PPV), Sensitivity/Recall, and Specificity. The F_1 score represents the harmonic mean of the Precision and Recall; NPV is a calculation of true negatives / all negatives; PPV is a calculation of true positives / all positives; Sensitivity/Recall is a calculation of true positives / (all true positives and false negatives); and Specificity is a calculation of true negatives / (all true negatives and false positives). TN, FN, TP, and FP refer to true negatives, false negatives, true positives, and false positives respectfully.

Metric	Model
Negative Predictive Value (NPV)	$\frac{\text{TN}}{\text{TN+FN}}$
Positive Predictive Value (PPV or Precision)	$\frac{\text{TP}}{\text{TP+FP}}$
Sensitivity (Recall)	$\frac{\text{TP}}{\text{TP+FN}}$
Specificity	$\frac{\text{TN}}{\text{TN+FP}}$
F_1 Score	$2 \times \frac{\text{precision} \times \text{recall}}{\text{precision+recall}}$

There is no accepted method of power calculation for elastic net models. Instead, we sought to assess the *post hoc* power to detect a pre-specified effect size for each predictor in the *Gene&Binary&Count* model. Our goal was to, at least roughly, assess whether the observational study could have detected a moderate effect of each predictor given the observed sample sizes and assuming no interaction with other predictors. Whether a predictor actually gets selected into the fitted elastic net model, however, depends on the complicated relationships between predictors [61], which was ignored in our power analysis. We based our analysis on simple logistic regression and a Wald test-derived procedure described in [61] and implemented in the R *WebPower* package [61, 62] with the `wp.logistic()` function. We set $\alpha = 0.05$ and distribution family to "Bernoulli" when testing binary predictors and to "normal" when testing the "count" predictor in relevant models. In all cases, we sought the power of detecting effect sizes of $\beta_1 = 0.7$, 1.1, 1.6, 2.3 or a 2, 3, 5 or 10-fold increase in the odds of resistance upon inclusion (or unit increase) of the predictor. For the binary predictor x_{ijk} indicating the presence of gene k conferring resistance to antibiotic i, we set the null model resistance probability (parameter `p0`) to the observed proportion in the subset of observations j with $x_{ijk} = 0$ and the alternative model resistance probability (parameter `p1`) was set to the expected proportion of resistance upon the desired fold increase in odds of resistance. For the "count" predictor, we set `p0` to the observed

proportion of resistance in the sample and `p1` to the proportion obtained upon increasing the odds of resistance by the desired fold.

Results

We examined the prevalence of antibiotic resistant or non-*wt* phenotypes across 980 *E. coli* isolates. When evaluating prevalence, we only consider phenotypes with at least five samples. The highest prevalence of resistance was for minocycline (100%; $n = 21$), doxycycline (94.5%; $n = 52$), and ampicillin (92.5%; $n = 571$). The lowest prevalence was for amikacin (6.45%; $n = 22$) and piperacillin-tazobactam (6.74%; $n = 13$). Non-*wt* phenotypes were most prevalent against imipenem (66.7%; $n = 4$), ampicillin (52.6%; $n = 91$) and amoxicillin (35%; $n = 57$). We observed zero isolates with non-*wt* phenotypes against amikacin ($n = 175$). The following antibiotics had non-interpretable MIC values; azithryomycin ($n = 116$), cephalexin ($n = 152$), and florfenicol ($n = 65$).

Previous studies have shown isolate resistance varies between different animal species [28, 29, 63–65], and we confirm isolates from distinct animal hosts exhibit widely varying rates of AMR. We used a Fisher's exact test to identify significant differences in the proportion of resistant isolates between host species for multiple antibiotics, with respect to CLSI or ECOFF values (Fig 2). For streptomycin and neomycin, the rate of resistance was significantly lower in chicken relative to isolates from other host animals. For gentamicin, another aminoglycoside, dog isolates had the lowest rate of resistance. Horse isolates had the largest differences in resistance against three drugs relative to other animal hosts, with over 85% of isolates resistant to cefazolin, ampicillin, and enrofloxacin. Cat, horse, and cattle isolates shared high rates of resistance to ampicillin, significantly greater than dog isolates. Although ECOFF values are not animal specific, we observe significant differences in the fraction of non-*wt* phenotypes for trimethoprim-sulphamethoxazole, chloramphenicol, tetracycline and doxycycline. Across shared antibiotics, turkey isolates had an overall higher proportion of resistant phenotypes than chicken isolates, and isolates from dogs had higher rates of resistance than isolates from cats. Even within similar groups of animals (e.g. avian species: chicken and turkey, or companion species: cats and dogs), there are significant differences in rates of resistance observed for a given antibiotic.

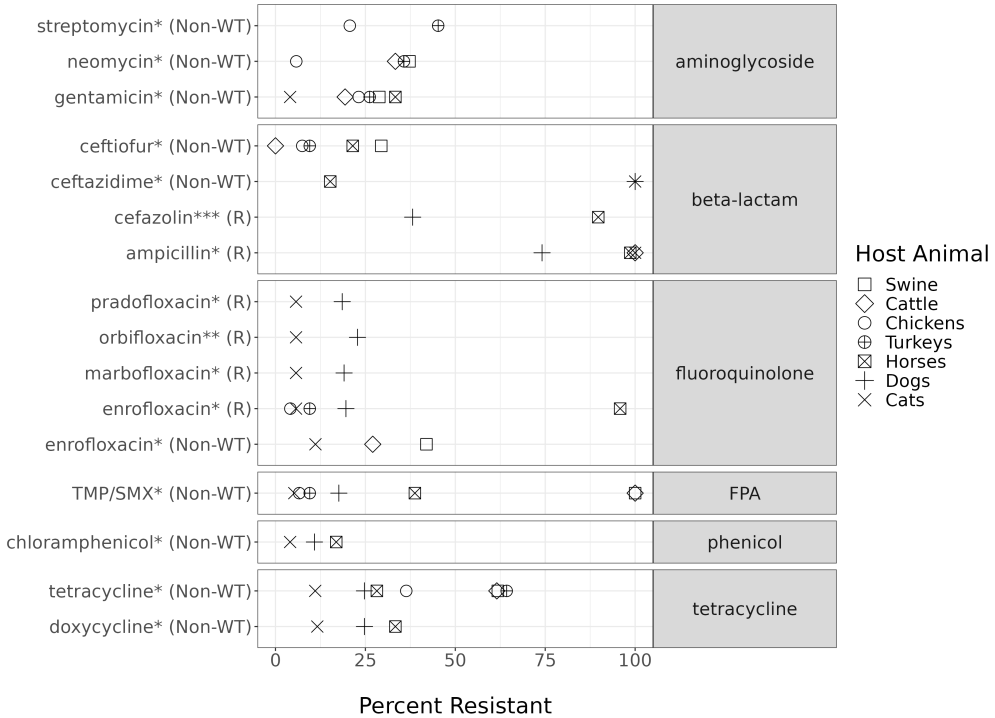

Fig 2. Significant differences in antibiotic resistance between animal hosts. The y-axis labels indicate the antibiotic and breakpoints used to determine resistant phenotype. The x-axis represents the percentage of resistant isolates. Symbols indicate animal host. Each point is the proportion of resistant isolates to the named antibiotic among those extracted from the indicated animal host. Only antibiotics with a significant difference in the proportion of resistance isolates between animal species are shown (adjusted p-value, $* \leq 0.05$, $** \leq 0.01$, $*** \leq 0.001$). FPA = Folate Pathway Antagonist, TMP/SMX = trimethoprim-sulphamethoxazole.

Predicting resistance

We fit increasingly complex elastic net models corresponding to types of association for the genetic determinants of AMR (Fig 3). We assume genotype-to-phenotype associations will vary depending on the host, so each model includes host animal main and interaction effects as a baseline. Many existing AMR models predict resistance based on the presence of *any* AMR genes relevant to the modeled antibiotic [66–68]. A relevant AMR gene is a gene known to confer resistance to a specific (or response) antibiotic. We fit an equivalent model, *Binary*, and measure its F_1 score, or the harmonic mean of the precision and recall at $69.2\% \pm 24.1\%$. We subsequently fit more complex models building upon reasonably assumed mechanisms of resistance. The *Count* model, which evaluates whether the log odds of resistance linearly increases with the presence of every additional AMR gene, achieved a mean F_1 of $92.0\% \pm 10.3\%$ across all antibiotics. The *Gene* model uses the presence of individual AMR genes as separate predictors, allowing different genes to have distinct roles in conferring resistance. The *Gene* model, despite having more predictors, only achieved a F_1 of $84.1\% \pm 18\%$, lower than the *Count* model.

We separately added the "binary" or "count" predictors to the *Gene* model, allowing the underpowered data sets to use simpler predictors in addition to individual genes. The *Gene&Binary* model achieved a slightly higher mean F_1 of $85.5\% \pm 16.1\%$. The *Gene&Count* achieved the best performance with a mean F_1 of $98.2\% \pm 2.3\%$, with a similar performance to that of the *Gene&Binary&Count* model with a mean F_1 of

98.3% ± 2.2%. While the performance of the *Gene&Count* and *Gene&Binary&Count* models were nearly equivalent (98.2% ± 2.3% and 98.3% ± 2.2%, respectively), the addition of the "binary" predictor to the *Gene* model conferred some information on the resistance relationship, although it is mostly eclipsed by the more pertinent and predictive information conferred by "count". It should be noted that on the training set, the *Gene&Binary&Count* model achieved a mean F_1 of 98.6% ± 2.6%. The improvement in performance over single mechanism models (*Binary*, *Count*, and *Gene*) confirms that AMR genes do not contribute equally to antibiotic resistance. We also fit a *Full* model that considered all identified AMR genes, not exclusively those expected to confer resistance to the model antibiotic, and found the additional predictors did not lead to improved performance over the *Gene&Count* combination (F_1 of 88.7% ± 14.4%).

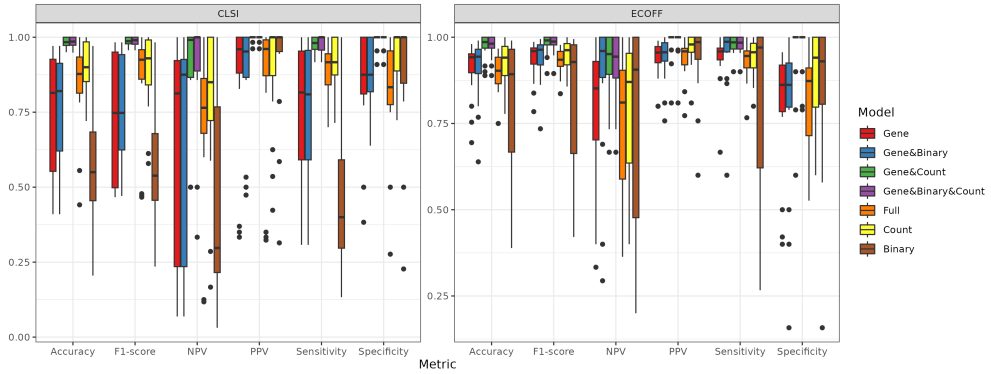

Fig 3. Elastic net model evaluation. We fit seven different elastic net models evaluating the effect of different predictors described in Table 1, including: AMR genes, animal host, presence of $n \geq 1$ AMR gene ("binary"), and total counts of relevant AMR genes (n, "count"). The *Gene&Binary&Count* model best predicts the resistance phenotype, with the highest median F_1 score and median accuracy across the different antibiotics. Model performance was evaluated using F_1 score, NPV, PPV, Sensitivity/Recall, and Specificity. NPV = Negative Predictive Value; PPV = Positive Predictive Value.

Across all evaluated antibiotics and breakpoints, the *Count* model (i.e. number of genes) demonstrated the strongest improvement in phenotype prediction compared with additional genes or the *Binary* model. The *Count* model achieved performance exceeding more complex models using only a single predictor. Surprisingly, the performance of the *Count* model suggests the number of resistance genes is more important for predicting resistance than specific gene identity, and individual genes alone cannot predict resistance. At the same time, it is also possible the *Count* model offers a low-dimensional approximation to a more complex resistance relationship involving multiple genes and their interactions that cannot be supported by the limited data in this study. Still, there were several antibiotics where the "count" variable was insufficient to predict resistance by itself. Both the *Count* and *Gene* models achieved F_1 scores ≤ 76.9 for amoxicillin/clavulanic acid, ampicillin, and cephalexin. Notably, the joint *Gene&Count* model achieved improved F_1 scores of ≤ 95.7 for the same antibiotics.

All models included animal main effects and animal×gene interaction effects to allow different levels and types of genetic associations across animal host. To evaluate the role of animal effects in these models, we removed animal main and interaction effects from the best performing *Gene&Binary&Count* model. The performance of the *Gene&Binary&Count* model without animal effects was nearly equivalent with an achieved mean F_1 of 98.0% ± 2.4%. Despite our initial assumptions, the inclusion of animal predictors did not impact resistance prediction on average. The lack of animal host influence on AMR resistance could be explained by the isolation of animal hosts from each other, thus limiting cross-species transmission of AMR genes. The co-linearity between different AMR genes and animal hosts would allow genetic

predictors to effectively substitute for animal effects [69], but a no-animal *Count* model, which masked the identity of individual genes and did not use any animal information, achieved only a slightly lower mean F_1 score of $90.5\% \pm 14.7\%$ compared with the animal *Count* model (mean F_1 of 92.0%).

Analysis of individual predictors

Because all prior results are averaged across models, it is important to understand the contribution of individual predictors to individual models and to help establish an interpretable biological understanding of AMR. We decided to explore the individual predictors in the *Gene+Binary+Count* model Although inclusion of the "binary" and animal predictors had little average effect, the elastic net handles the excess predictors with no decrease in average performance, allowing assessment of the importance of all these predictors in the individual models. We evaluated the importance, effect size, and power to detect each predictor. Given the high number of possible genetic predictors and the inevitable imbalance in predictors and response, this observational study is under-powered for many possible predictors. We therefore computed for each predictor a power to detect with a simplified model (see Methods). Across all 31 models, the average power to detect any predictor with effect size of 0.7 (two-fold increase in odds of resistance) at significance level 0.05 in the *Gene&Binary&Count* model was only 0.56 (sd = 0.22%). At an effect size of 2.3, equivalent to a ten-fold increase in the odds of resistance, the same average power was 0.90 (sd = 0.28%).

Next, we wanted estimate the relative contribution of the host animal effect. To do so we calculated the proportion of important animal-related predictors, which include the animal host predictors and their interactions with AMR genes, for each model. Of the 31 fitted models, 14 did not use any animal-related predictors, and we observed that models with fewer predictors do not consider animal information. Models with 30 or more predictors had a majority of animal-related predictors. The non-*wt* doxycycline, chloramphenicol, and neomycin were the only models with less than 27 total predictors that did not use any animal information. The models with a high proportion ($\geq 80\%$) of animal interaction terms among their important predictors were non-*wt* trimethoprim-sulphamethoxazole (93.2%), cefazolin resistance (86.9%), ampicillin resistance (84.9%), non-*wt* ceftazidime (83.6%), non-*wt* amoxicillin (82.9%), gentamicin resistance (82.9%), non-*wt* spectinomycin (82%), and non-*wt* streptomycin (80%). There was a trend for non-*wt* models to select animal predictors, with 12/16 (75%) incorporating animal information, while 10/15 (67%) of resistance models did not use any. While some models used animal information, the total importance of animal predictors in the *Gene&Binary&Count* model was relatively low (Table S3), consistent with the already reported small improvement in average model performance.

Unique predictors for resistance within an antibiotic class.

Next we sought to understand the relative complexity of AMR genotype-to-phenotype relationships across different antibiotics. To do so, we compared the number of important predictors for each model, by comparing the number of important predictors was correlated with the number of candidate predictors for each antibiotic (Pearson's correlation, $R = 0.61$). The model for amikacin and cefovecin selected a large proportion, 30/35 (85.7%) and 84/98 (85.7%), respectively, of their candidate predictors as important (Table 3). However, this pattern was not observed in other β-lactam class antibiotics. For ampicillin resistance, important predictors were only 26.7% of candidates, and amoxicillin-clavulanic acid resistance only selected 8/95 (8.4%) of possible predictors. The fraction of important predictors was similar for antibiotics

across CLSI and ECOFF breakpoints, the model for non-*wt* ampicillin selected 12 important predictors out of 194 or 6.1%.

In contrast with the varying complexity of β-lactam models, the models for fluoroquinolones were much simpler. Despite good power to detect a role for individual genes, only the "count" predictor was consistently selected for both resistant and non-*wt* phenotype prediction. For marbofloxacin, orbifloxacin, and pradofloxacin resistance models, this may be caused by the relatively low number of candidate predictors ($n = 11$). However, the models for resistant and non-*wt* enrofloxacin have more candidate predictors comparable with several β-lactams yet still only select for the "count" predictor. The sharing of a single predictor across different models suggests resistance or non-*wt* phenotype prediction is more similar for fluoroquinolones than other antibiotics. Aminoglycoside resistance shared a large fraction of important predictors, 22/29 (73.3%) for amikacin and 22/41 (53.6%) for gentamicin, respectively. These similarities were not present for non-*wt* aminoglycoside phenotype prediction, where the important predictors for gentamicin, neomycin, and streptomycin were mostly unique. A table of important model predictors for each antibiotic and breakpoint, not including animal interaction predictors, is provided in the supplement (Table S4).

Table 3. Unique predictors in fitted *Gene&Binary&Count* elastic net model by antibiotic. The important predictors included in the fitted *Gene&Binary&Count* models for each antibiotic were identified and compared to the shared important predictors for antibiotics within the same class. For each antibiotic class, "Shared" is the number of important predictors selected in all antibiotics within the same class, the denominator indicates the total number shared predictors. $|\hat{\mathcal{X}}_i|$ is the total number of important predictors selected by the model. $\sum_{i=1}^{j} \beta_i$ is the sum of coefficient absolute values for predictors. $|\mathcal{X}_i|$ refers to the number of candidate predictors considered. $|\mathcal{X}_i^{'}|$ refers to the number of candidate predictors with estimated power ≥ 0.80 at an effect size of 2.3. "Unique" is the number of important predictors $|\mathcal{X}_i|$ that are unique to that antibiotic relative to its class. β-lactams and β-lactam combination antibiotics had the highest number of total and unique important predictors spread across each of the antibiotics in its class. Models for fluoroquinolone resistance only agreed on a single predictor (*Count*).

BP	Class	Antibiotic	Shared	Unique	$\lvert\hat{\mathcal{X}}_i\rvert$	$\sum_{i=1}^{j} \beta_i$	$\lvert\mathcal{X}_i^{'}\rvert$	$\lvert\mathcal{X}_i\rvert$
CLSI	aminoglycoside	amikacin	22/29	8	30	37.6	35	35
		gentamicin		19	41	50	53	53
	β-lactam	TZP[a]	2/17	6	27	190	34	71
		ampicillin		22	73	81.4	273	274
		cefazolin		43	99	247	127	128
		cefovecin		37	84	53.3	97	98
		cefpodoxime			16	115	34	71
		ceftazidime			8	26.3	24	51
		ceftiofur		1	22	148	33	71
		cephalexin			15	85	26	55
		co-amoxiclav[b]			8	23.7	94	95
	fluoroquinolone	enrofloxacin	1/11		1	0.476	41	41
		marbofloxacin			1	0.258	10	11
		orbifloxacin			1	0.44	11	11
		pradofloxacin			1	0.258	10	11
ECOFF	FPA	TMP/SMX[c]		88	88	43.7	127	127
	aminocyclitol	spectinomycin	3/13	50	50	62	89	89
	aminoglycoside	gentamicin	2/17	36	41	110	146	146
		neomycin		15	24	56.5	64	64
		streptomycin		28	40	94.9	47	47
	β-lactam	amoxicillin		65	82	76.1	100	101
		ampicillin			12	0.962	193	194
		cefazolin		31	65	43.1	54	83
		cefpodoxime		1	15	215	25	53
		ceftazidime		9	55	31.1	182	183
		ceftiofur		28	85	122	346	347
		co-ticarclav[d]		2	12	94.9	32	69
	fluoroquinolone	enrofloxacin		1	1	0.39	35	35
	phenicol	chloramphenicol		22	22	40.4	51	51
	tetracycline	doxycycline	5/27	6	11	35	27	27
		tetracycline		34	39	70.6	79	79

[a] piperacillin-tazobactam [b] amoxicillin-clavulanic acid [c] trimethoprim-sulphamethoxazole
[d] ticarcillin-clavulanic acid

The difference in important predictors for resistance to antibiotics within the same class may indicate that differences do exist in the genetic associations for resistance. This difference could also be explained by insufficient data in a particular antibiotic and breakpoint subset to study the antibiotic or low power when considering smaller effect sizes.

Discussion

The availability of sequence data has been proposed as a major tool for monitoring antibiotic resistance [70–72]. Multiple studies have used WGS to predict resistance phenotypes in pathogens such as *E. coli*, *Klebsiella pneumoniae*, *Salmonella enterica* serovar Typhimurium, and *Staphylococcus aureus* [73–75]. These studies report high concordance between a pathogen's AMR profile and predicted resistance phenotype, and consequently affirm the potential of sequencing methods to improve AMR surveillance. In recent years, the predictive accuracy the ability to extract useful information about these relationships has increased with the development of predictive ML models for AMR [13, 76]. However, we found these AMR genotype-phenotype relationships to vary substantially depending on the animal host and antibiotic, indicating a degree of AMR-predictive complexity that has not yet been discussed. These results are supported by earlier studies which show varying rates of resistance for antibiotics across animal hosts [28, 29]. While the potential utility and use of bacterial genotype data in monitoring AMR is certain, there is a need to refine current approaches to account for the biological variation in AMR mechanisms between different host animals, antibiotics, and known AMR conferring genes.

AMR genotype-phenotype relationships are biologically complex.

We observe the largest improvement in F_1 score from the *Gene* model to *Gene&Count* models in cephalexin, ampicillin, amoxicillin-clavulanic acid, and fluoroquinolone class antibiotic resistance. This suggests the additive effect of multiple AMR genes against those antibiotics is a more effective predictor of resistance than the identity of the genes involved. From a biomechanistic perspective, orbifloxacin, a fluoroquinolone, disrupts bacterial cells by interfering with DNA replication enzymes such as topoisomerase and DNA gyrase. One mechanism of fluoroquinolone resistance is the synthesis of proteins to competitively inhibit DNA gyrase and prevent fluoroquinolones from disrupting its function [77]. For these antibiotics, the number of genes and, in turn, the number of synthesized proteins, may be a more accurate predictor of resistance than the identity of individual genes. In which case, transcriptomic data may be invaluable for fluoroquinoline resistance prediction. Another possibility is that the *Gene* model is over-parameterized in these data, and a more parsimonious model with a single predictor could achieve better performance, but in most cases the power to detect individual genes in quite high. Controlled experiments could confirm the importance of the "count" predictor, or additional samples may yet yield the statistical power to reveal the importance of gene identities. Furthermore, while the "count" of relevant AMR genes is an important predictor of resistance and models including the "count" predictor exceed the performance of the rule-based model (*Binary*) for certain antibiotics, the more complex joint model using the genomic, animal host, count, and binary predictors (*Gene&Binary&Count*) performs better still.

The genetic complexity of antibiotic resistance, in terms of the number of important predictors, was much higher for β-lactam and β-lactam combination antibiotics. The β-lactam class models, using both CLSI and ECOFF values, selected the highest number of unique predictors per antibiotic. More specifically, the most complex resistance models were for cefazolin, cefovecin and ampicillin. Although many of the important predictors in these models correspond to subclasses of the β-lactamase gene, their significance as unique predictors suggests they are non-substitutive and have varying importance in conferring resistance. Differences in β-lactamase protein target affinity and degree of membrane permeability may make certain subclasses of β-lactamase proteins more suited to act on specific antibiotics [78]. The model for resistance to amoxicillin-clavulanic acid, a different β-lactam class antibiotic, only

selected for eight important predictors. This difference in complexity may be due to unique properties of amoxicillin-clavulanic acid or genetic mechanisms unique to amoxicillin-clavulanic acid resistance.

The difference in model complexity for each antibiotic type suggests there are antibiotic-specific AMR genotype-to-phenotype complex relationships more complex than have been previously explored [23–25]. While WGS information can be used to accurately predict AMR phenotype, the most effective way to use that information may differ between antibiotics. As an example, there were significant differences in the odds of resistance between different animal hosts for some antibiotics, and animal predictors were used to improve the predictive power of our models. Yet for other antibiotics, such as fluoroquinolones, our results show animal predictors were not necessary to make accurate predictions. In addition, the choice of specific important predictor selection in our model was significantly affected by the power to detect those predictors. Animal-specific fluoroquinolone interactions may affect resistance at a lower effect size than what is detectable in our study.

Notably, we observed a difference in the importance of selection for animal predictors in non-*wt* and resistant phenotypes. Animal predictors and interactions are more commonly found in models predicting resistance, suggesting that the animal/tissue differences respected by CLSI breakpoints also associate with differences in AMR genetic determinants. As expected, ECOFF-determined *wt* phenotypes, which are host independent, yield fewer important animal predictors. On the other hand, the development of CLSI breakpoints do take into consideration the presence/absence of numerous clinical factors, including animal host species, source tissue, and, in turn, antimicrobial- and animal-specific pharmacokinetics.

Our observations suggest that using the elastic net approach with animal predictors can accurately predict antibiotic non-*wt* phenotypes, even when those phenotypes are established using inherently non-clinical ECOFF values. As a result, the work shown here would suggest that specific antibiotics: namely, chloramphenicol, doxycycline, tetracycline, trimethoprim sulfamthoxazole, amoxicillin, neomycin, and spectinomycin, be prioritized for CLSI breakpoint analysis and development.

AMR prediction is accurate, but incomplete.

Both the *Gene&Count* and *Gene&Binary&Count* models were able to accurately predict isolate resistance. Although performance varied between different antibiotics, mean the F_1 for both models were very high (0.988). All models share a higher average positive predictive value than negative predictive value, suggesting there are types of AMR which are not accounted for in the model and would improve negative prediction. This same complexity may also increase the chance that important resistance-conferring genes are missing from the tested predictors [79]. Theoretical power calculations for elastic net models have yet to be defined, but calculations based on traditional logistic regression suggest that despite the size of this study, our approach was underpowered considering: (1) the number of predictors and (2) an effect size equivalent to a two-fold increase in the likelihood of resistance, even when limited to only the relevant AMR genes for each antibiotic. Therefore, important predictors of AMR would best be assessed in an experiment explicitly designed to detect the effect of multiple genetic predictors. This type of experiment would require significantly more samples than were observed here to ensure sufficient statistical power, especially when measuring smaller effect sizes.

Conclusions

Our study illustrates the genetic and environmental complexity of AMR that is often ignored in models based on ML. We confirmed that existing genetic determinants of AMR are generally accurate predictors of antibiotic resistance across animal hosts and antibiotics. However, we find evidence for possible differences in the mechanisms of AMR across host species. Using elastic net models, which offer more interpretable features than other high dimensional models, we find the genotype/phenotype relationship for AMR depends on multiple factors, including the genetic association with AMR, animal host, antibiotic, and AMR genes. While our study identifies strong genetic predictors of AMR resistance from observational samples, these predictors are merely a subset of the candidate predictor genes that we expect to confer resistance. It is uncertain if predictors not included in the model are unnecessary for predicting resistance, or whether we lack sufficient power to measure their effect on resistance. We suggest these results be further elucidated in a more powerful study, where smaller effect sizes can be detected. the effect size of individual genes can more accurately be measured. It might also be possible to test some of the biological mechanisms revealed from our interpretable elastic net models. While there is considerable evidence that WGS data are highly predictive of AMR phenotype and will improve AMR prevention strategies, we suggest there is still much to learn about the genotype-to-phenotype relationships and their differences across hosts and antibiotics. Models fit to data from multiple hosts and antibiotics may be most useful when there are commonalities in resistance mechanisms across hosts and antibiotics (e.g. in the same class), but our results show such models must also allow for some differences. In the end, such models may improve early diagnosis across hosts, and help us take action to combat resistance, especially when resistance patterns that also threaten human health emerge in non-human hosts.

Author Declarations

Data Availability Statement

Raw sequences used to produce the AMR output data are publicly available on NCBI's Sequence Read Archive under BioProject ID PRJNA510384 (`https://www.ncbi.nlm.nih.gov/bioproject/510384`). The code and data are available at `https://www.github.com/FriedbergLab/AMREcoli` and on Figshare at `https://doi.org/10.6084/m9.figshare.21737288.v1`.

Author Contributions

HCC, CLF, IF, KSD, and BH designed the research plan and experimental schema; HCC, CLF, JAH, TS, and BH conducted the experiments and acquired data for research analysis; all authors analyzed and interpreted the data. All authors also drafted, critically revised, and reviewed the manuscript for important intellectual content and agree to be held accountable to the accuracy and integrity of all work represented here.

Acknowledgments

We gratefully acknowledge Dr. Christina Loiacono, Dr. Kelley Black, Dr. Kristina Lantz, Keira Stuart, Kathryn Reynolds, and Jennifer Rodriguez in the NVSL for their valuable contributions to data aggregation, collection, and processing across the NAHLN participating laboratories over the past five years. We also gratefully acknowledge the NAHLN participating laboratories that contributed WGS and AST data to this project:

Auburn University Bacteriology and Mycology Diagnostic Laboratory, Auburn, AL
California Animal Health and Food Safety Laboratory System, Davis, CA
Colorado State University Veterinary Diagnostic Laboratory, Fort Collins, CO
Kissimmee Animal Disease Diagnostic Laboratory, Kissimmee, FL
Athens Veterinary Diagnostic Laboratory, Athens, GA
Iowa State University Veterinary Diagnostic Laboratory, Ames, IA
Purdue University Animal Disease Diagnostic Laboratory, West Lafayette, IN
Kansas State University Veterinary Diagnostic Laboratory, Manhattan, KS
Murray State University, Breathitt Veterinary Center, Hopkinsville, KY
University of Kentucky Veterinary Diagnostic Laboratory, Lexington, KY
Louisiana State University Animal Disease Diagnostic Laboratory, Baton Rouge, LA
Michigan State Diagnostic Center for Population and Animal Health, Lansing, MI
University of Minnesota Veterinary Diagnostic Laboratory, MN
University of Missouri Veterinary Medical Diagnostic Laboratory, Columbia, MO
Mississippi State University Veterinary Research and Diagnostic Laboratory, Pearl, MS
North Carolina Veterinary Diagnostic Laboratory System, Rollins Animal Disease Diagnostic Laboratory, Raleigh, NC
North Dakota State University Veterinary Diagnostic Laboratory, Fargo, ND
University of Nebraska Veterinary Diagnostic Center, Lincoln, NE
Cornell Animal Health Diagnostic Center, Ithaca, NY
Ohio Department of Agriculture, Reynoldsburg, OH
Pennsylvania Animal Diagnostic Laboratory System, University of Pennsylvania New Bolton Center, Kennett Square, PA
Pennsylvania Animal Diagnostic Laboratory System, Pennsylvania Veterinary Laboratory, Harrisburg, PA
Pennsylvania State University Animal Diagnostic Laboratory, University Park, PA
South Dakota State University Animal Disease Research and Diagnostic Laboratory, Brookings, SD
Texas A&M University Veterinary Medical Diagnostic Laboratory, College Station, TX
Washington State University Animal Disease Diagnostic Laboratory, Pullman, WA
University of Wisconsin Veterinary Diagnostic Laboratory, Madison, WI

Disclaimer

The views expressed in this article are those of the authors and do not necessarily reflect the policies and views of the USDA, DOE, or ORAU/ORISE. Mention of trade names or commercial products in this publication are solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

Abbreviations

AAVLD, American Association of Veterinary Laboratory Diagnosticians
AMR, antibiotic resistance
APHIS, USDA Animal and Plant Health Inspection Service
AST, antibiotic sensitivity testing
BHI, brain heart infusion
CARB NAP, National Action Plan for Combating Antibiotic-Resistant Bacteria
CEAH, USDA Center for Epidemiology and Animal Health
CDC, Centers for Disease Control and Prevention
co-amoxiclav, amoxicillin-clavulanic acid
co-ticarclav, ticarcillin-clavulanic acid
CLSI, Clinical Laboratory Sciences

ECOFF, epidemiological cut-off
EUCAST, European Committee on Antimicrobial Susceptibility
FDA, U.S. Food and Drug Administration
FDA-CVM-VetLIRN, FDA Center for Veterinary Medicine Veterinary Laboratory Investigation and Response Network
FPA, folate pathway antagonist
MALDI-TOF, matrix-assisted laser desorption/ionization–time of flight
MIC, minimum inhibitory concentration
ML, machine learning
NAHLN, National Animal Health Laboratory Network
NARMS, National Antibiotic Resistance Monitoring System
NCBI, National Center for Biotechnology Information
NPV, negative predictive value
NVSL, National Veterinary Services Laboratories
ORISE, Oak Ridge Institute for Science and Education
ORAU, Office of Research in Associated Universities
PPV, positive predictive value
SBA, sheep blood agar
TMP/SMX, trimethoprime-sulphamethoxazole
TrAC, Translational Artificial Intelligence Center
TSB, trypticase soy broth
TZP, piperacillin-tazobactam
USDA, U.S. Department of Agriculture
WGS, whole-genome sequencing

References

1. Organization WH. Global Action Plan on Antimicrobial Resistance; World Health Organization: Geneva, Switzerland. Geneva, Switzerland: World Health Organization; 2016.

2. Elta S, Catherine A L, Jirka T, Calum M, Louise L, Emma H, et al. Evaluation of the Action Plan against the Rising Threats from Antimicrobial Resistance. Brussels, Belgium: European Comission; 2016.

3. United States Center for Disease Control. Antibiotic Resistance Threats in the United States, 2019. United States Center for Disease Control; 2019.

4. Department of Health and Human Services. Antibiotic resistance threats in the United States, 2013; 2013.

5. United States Food and Drug Administration. National action plan for combating antimicrobial resistant bacteria; 2015. Available from: `https://obamawhitehouse.archives.gov/sites/default/files/docs/national_action_plan_for_combating_antibotic-resistant_bacteria.pdf`.

6. White A, Hughes JM. Critical Importance of a One Health Approach to Antimicrobial Resistance. EcoHealth. 2019;16(3):404–409. doi:10.1007/s10393-019-01415-5.

7. Karp BE, Tate H, Plumblee JR, Dessai U, Whichard JM, Thacker EL, et al. National Antimicrobial Resistance Monitoring System: Two Decades of Advancing Public Health Through Integrated Surveillance of Antimicrobial Resistance. Foodborne Pathogens and Disease. 2017;14(10):545–557. doi:10.1089/fpd.2017.2283.

8. U S Food and Drug Administration. Timeline of FDA action on antimicrobial resistance; 2022. Available from: `https://www.fda.gov/animal-veterinary/antimicrobial-resistance/timeline-fda-action-antimicrobial-resistance`.

9. Animal and Plant Health Inspection Service. National Animal Health Laboratory Network (NAHLN) Antimicrobial Resistance Pilot Project Year 1 report: 2018. United States Food and Drug Administration; 2018.

10. Boolchandani M, D'Souza AW, Dantas G. Sequencing-based methods and resources to study antimicrobial resistance. Nature Reviews Genetics. 2019;20:356–370. doi:10.1038/s41576-019-0108-4.

11. McArthur AG, Tsang KK. Antimicrobial resistance surveillance in the genomic age. Ann N Y Acad Sci. 2017;1388(1):78–91.

12. Bhattacharyya RP, Bandyopadhyay N, Ma P, Son SS, Liu J, He LL, et al. Simultaneous detection of genotype and phenotype enables rapid and accurate antibiotic susceptibility determination. Nature Medicine. 2019;25(12):1858–1864.

13. Anahtar MN, Yang JH, Kanjilal S. Applications of Machine Learning to the Problem of Antimicrobial Resistance: an Emerging Model for Translational Research. Journal of Clinical Microbiology. 2021;59(7):e0126020–e0126020. doi:https://doi.org/10.1128/JCM.01260-20.

14. Niehaus KE, Walker TM, Crook DW, Peto TEA, Clifton DA. Machine learning for the prediction of antibacterial susceptibility in Mycobacterium tuberculosis. In: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI). IEEE; 2014.Available from: https://doi.org/10.1109/bhi.2014.6864440.

15. Pesesky MW, Hussain T, Wallace M, Patel S, Andleeb S, Burnham CAD, et al. Evaluation of Machine Learning and Rules-Based Approaches for Predicting Antimicrobial Resistance Profiles in Gram-negative Bacilli from Whole Genome Sequence Data. Frontiers in Microbiology. 2016;7. doi:10.3389/fmicb.2016.01887.

16. Eyre DW, Silva DD, Cole K, Peters J, Cole MJ, Grad YH, et al. WGS to predict antibiotic MICs for Neisseria gonorrhoeae. Journal of Antimicrobial Chemotherapy. 2017;72(7):1937–1947. doi:10.1093/jac/dkx067.

17. Yang Y, Niehaus KE, Walker TM, Iqbal Z, Walker AS, Wilson DJ, et al. Machine learning for classifying tuberculosis drug-resistance from DNA sequencing data. Bioinformatics. 2017;34(10):1666–1671. doi:10.1093/bioinformatics/btx801.

18. Yang Y, Walker TM, Walker AS, Wilson DJ, Peto TEA, Crook DW, et al. DeepAMR for predicting co-occurrent resistance of Mycobacterium tuberculosis. Bioinformatics. 2019;35(18):3240–3249. doi:10.1093/bioinformatics/btz067.

19. Nguyen M, Long SW, McDermott PF, Olsen RJ, Olson R, Stevens RL, et al. Using Machine Learning To Predict Antimicrobial MICs and Associated Genomic Features for Nontyphoidal Salmonella. Journal of Clinical Microbiology. 2019;57(2). doi:10.1128/jcm.01260-18.

20. Maguire F, Rehman MA, Carrillo C, Diarra MS, Beiko RG. Identification of Primary Antimicrobial Resistance Drivers in Agricultural Nontyphoidal Salmonella enterica Serovars by Using Machine Learning. mSystems. 2019;4(4). doi:10.1128/msystems.00211-19.

21. Pearcy N, Hu Y, Baker M, Maciel-Guerra A, Xue N, Wang W, et al. Genome-Scale Metabolic Models and Machine Learning Reveal Genetic Determinants of Antibiotic Resistance in Escherichia coli and Unravel the Underlying Metabolic Adaptation Mechanisms. mSystems. 2021;6(4). doi:10.1128/msystems.00913-20.

22. McDermott PF, Davis JJ. Predicting antimicrobial susceptibility from the bacterial genome: A new paradigm for one health resistance monitoring. J Vet Pharmacol Ther. 2021;44(2):223–237.

23. Tyson GH, McDermott PF, Li C, Chen Y, Tadesse DA, Mukherjee S, et al. WGS accurately predicts antimicrobial resistance in Escherichia coli. Journal of Antimicrobial Chemotherapy. 2015;70(10):2763–2769. doi:10.1093/jac/dkv186.

24. McDermott PF, Tyson GH, Kabera C, Chen Y, Li C, Folster JP, et al. Whole-Genome Sequencing for Detecting Antimicrobial Resistance in Nontyphoidal Salmonella. Antimicrobial agents and chemotherapy. 2016;60(9):5515–5520. doi:https://doi.org/10.1128/AAC.01030-16.

25. Zhao S, Tyson GH, Chen Y, Li C, Mukherjee S, Young S, et al. Whole-Genome Sequencing Analysis Accurately Predicts Antimicrobial Resistance Phenotypes in Campylobacter spp. Applied and environmental microbiology. 2016;82(2):459–466. doi:10.1128/AEM.02873-15.

26. Institute of Medicine and National Research Council, editor. The Use of Drugs in Food Animals. National Academies Press; 1999. Available from: https://doi.org/10.17226/5137.

27. Roth N, Käsbohrer A, Mayrhofer S, Zitz U, Hofacre C, Domig KJ. The application of antibiotics in broiler production and the resulting antibiotic resistance in Escherichia coli: A global overview. Poultry Science. 2019;98(4):1791–1804. doi:https://doi.org/10.3382/ps/pey539.

28. Garcia-Migura L, Hendriksen RS, Fraile L, Aarestrup FM. Antimicrobial resistance of zoonotic and commensal bacteria in Europe: The missing link between consumption and resistance in veterinary medicine. Veterinary Microbiology. 2014;170(1):1–9. doi:https://doi.org/10.1016/j.vetmic.2014.01.013.

29. Abdelwahab G, Ishag HZ, Hammadi Z, Yammahi S, Yusof M, Yassi M, et al. Antibiotics Resistance in Escherichia coli Isolated from from Livestock in the Emirate of Abu Dhabi, UAE, 2014–2019. International Journal of Microbiology. 2022;2022. doi:10.1155/2022/3411560.

30. Mellor KC, Petrovska L, Thomson NR, Harris K, Reid SWJ, Mather AE. Antimicrobial Resistance Diversity Suggestive of Distinct Salmonella Typhimurium Sources or Selective Pressures in Food-Production Animals. Frontiers in Microbiology. 2019;10. doi:10.3389/fmicb.2019.00708.

31. Leekitcharoenphon P, Johansson MHK, Munk P, Malorny B, Skarżyńska M, Wadepohl K, et al. Genomic evolution of antimicrobial resistance in Escherichia coli. Scientific Reports. 2021;11(1). doi:10.1038/s41598-021-93970-7.

32. Wassenaar TM, Silley P. Antimicrobial resistance in zoonotic bacteria: lessons learned from host-specific pathogens. Animal Health Research Reviews. 2008;9(2):177–186. doi:10.1017/s1466252308001539.

33. Furness LE, Campbell A, Zhang L, Gaze WH, McDonald RA. Wild small mammals as sentinels for the environmental transmission of antimicrobial resistance. Environmental Research. 2017;154:28–34. doi:10.1016/j.envres.2016.12.014.

34. Richards VP, Lang P, Bitar PDP, Lefébure T, Schukken YH, Zadoks RN, et al. Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted Streptococcus agalactiae. Infection, Genetics and Evolution. 2011;11(6):1263–1275. doi:10.1016/j.meegid.2011.04.019.

35. Hicks AL, Wheeler N, Sánchez-Busó L, Rakeman JL, Harris SR, Grad YH. Evaluation of parameters affecting performance and reliability of machine learning-based antibiotic susceptibility testing from whole genome sequencing data. PLoS Computational Biology. 2019;15(9):e1007349.

36. Osman KM, Kappell AD, Elhadidy M, ElMougy F, El-Ghany WAA, Orabi A, et al. Poultry hatcheries as potential reservoirs for antimicrobial-resistant Escherichia coli: A risk to public health and food safety. Sci Rep. 2018;8(1):5859.

37. Oniciuc EA, Likotrafiti E, Alvarez-Molina A, Prieto M, Santos JA, Alvarez-Ordóñez A. The present and future of whole genome sequencing (WGS) and whole metagenome sequencing (WMS) for surveillance of antimicrobial resistant microorganisms and antimicrobial resistance genes across the food chain. Genes (Basel). 2018;9(5).

38. Poirel L, Madec JY, Lupo A, Schink AK, Kieffer N, Nordmann P, et al. Antimicrobial Resistance in Escherichia coli. Microbiology spectrum. 2018;6(4). doi:10.1128/microbiolspec.ARBA-0026-2017.

39. Singhal N, Kumar M, Kanaujia PK, Virdi JS. MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. Frontiers in Microbiology. 2015;6. doi:10.3389/fmicb.2015.00791.

40. Clinical, Institute LS. Performance standards for antimicrobial disk and dilution susceptibility tests for bacteria isolated from animals. 4th ed. CLSI supplement, VET08 v. 38, no. 14.. Wayne, PA: Clinical and Laboratory Standards Institute; 2018.

41. Turnidge J, Kahlmeter G, Kronvall G. Statistical characterisation of bacterial wild-type MIC value distributions and the determination of epidemiological cut-off values. Clinical Microbiology and Infection. 2006;12(5):418–425. doi:10.1111/j.1469-0691.2006.01377.x.

42. Kahlmeter G. European harmonization of MIC breakpoints for antimicrobial susceptibility testing of bacteria. Journal of Antimicrobial Chemotherapy. 2003;52(2):145–148. doi:10.1093/jac/dkg312.

43. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Research. 2016;44(14):6614–6624. doi:10.1093/nar/gkw569.

44. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. PeerJ Computer science. 2017;3:e104–. doi:10.7717/peerj-cs.104.

45. Lu J, Rincon N, Wood DE, Breitwieser FP, Pockrandt C, Langmead B, et al. Metagenome analysis using the Kraken software suite. Nature Protocols. 2022;17(12):2815–2839. doi:10.1038/s41596-022-00738-y.

46. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, et al. Assembling Single-Cell Genomes and Mini-Metagenomes From Chimeric MDA Products. Journal of Computational Biology. 2013;20(10):714–737. doi:10.1089/cmb.2013.0084.

47. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, et al. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. Antimicrobial agents and chemotherapy. 2019;63(11):e00483–19. doi:https://doi.org/10.1093/ps/84.4.634.

48. Seemann T. Abricate; 2014. `https://github.com/tseemann/abricate`.

49. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. Journal of antimicrobial chemotherapy. 2012;67(11):2640–2644. doi:10.1093/jac/dks261.

50. R Core Team. R: A Language and Environment for Statistical Computing; 2020. Available from: `https://www.R-project.org/`.

51. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016. Available from: `https://ggplot2.tidyverse.org`.

52. Kuhn M, Wickham H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles.; 2020. Available from: `https://www.tidymodels.org`.

53. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B (Methodological). 1996;58(1):267–288.

54. Hoerl AE, Kennard RW. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics. 2000;42(1):80–86.

55. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology). 2005;67(2):301–320. doi:10.1111/j.1467-9868.2005.00503.x.

56. Kim JI, Maguire F, Tsang KK, Gouliouris T, Peacock SJ, McAllister TA, et al. Machine Learning for Antimicrobial Resistance Prediction: Current Practice, Limitations, and Clinical Perspective. Clinical Microbiology Reviews. 2022;35(3). doi:10.1128/cmr.00179-21.

57. Leng C, Lin Y, Wahba G. A note on the lasso and related procedures in model selection. Statistica Sinica. 2006; p. 1273–1284.

58. Ambroise C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. Proceedings of the National Academy of Sciences. 2002;99(10):6562–6566. doi:10.1073/pnas.102102699.

59. Khan MHR, Bhadra A, Howlader T. Stability selection for lasso, ridge and elastic net implemented with AFT models. Statistical Applications in Genetics and Molecular Biology. 2019;18(5). doi:10.1515/sagmb-2017-0001.

60. Meinshausen N, Bühlmann P. Stability selection. Journal of the Royal Statistical Society Series B (Methodological). 2010;72(4):417–473.

61. Demidenko E. Sample size determination for logistic regression revisited. Statistics in Medicine. 2007;26(18):3385–3397. doi:10.1002/sim.2771.

62. Zhang Z, Yuan KH. Practical statistical power analysis using webpower and R. ISDSA Press; 2018.

63. Lanz R, Kuhnert P, Boerlin P. Antimicrobial resistance and resistance gene determinants in clinical Escherichia coli from different animal species in Switzerland. Veterinary Microbiology. 2003;91(1):73–84. doi:https://doi.org/10.1016/S0378-1135(02)00263-8.

64. Bourély C, Coeffic T, Caillon J, Thibaut S, Cazeau G, Jouy E, et al. Trends in antimicrobial resistance among Escherichia coli from defined infections in humans and animals. Journal of Antimicrobial Chemotherapy. 2020;75(6):1525–1529. doi:10.1093/jac/dkaa022.

65. De Jong A, El Garch F, Hocquet D, Prenger-Berninghoff E, Dewulf J, Migura-Garcia L, et al. European-wide antimicrobial resistance monitoring in commensal *Escherichia coli* isolated from healthy food animals between 2004 and 2018. Journal of Antimicrobial Chemotherapy. 2022;77(12):3301–3311. doi:10.1093/jac/dkac318.

66. Tsang KK, Maguire F, Zubyk HL, Chou S, Edalatmand A, Wright GD, et al. Identifying novel β-lactamase substrate activity through in silico prediction of antimicrobial resistance. Microbial Genomics. 2021;7(1):000500. doi:10.1099/mgen.0.000500.

67. Tyson GH, Ceric O, Guag J, Nemser S, Borenstein S, Slavic D, et al. Genomics accurately predicts antimicrobial resistance in *Staphylococcus pseudintermedius* collected as part of Vet-LIRN resistance monitoring. Veterinary Microbiology. 2021;254:109006. doi:10.1016/j.vetmic.2021.109006.

68. Demczuk W, Martin I, Griffith A, Lefebvre B, McGeer A, Tyrrell GJ, et al. Linear Regression Equations To Predict β-Lactam, Macrolide, Lincosamide, and Fluoroquinolone MICs from Molecular Antimicrobial Resistance Determinants in *Streptococcus pneumoniae*. Antimicrobial Agents and Chemotherapy. 2022;66(1):e01370–21. doi:10.1128/AAC.01370-21.

69. Lees JA, Mai TT, Galardini M, Wheeler NE, Horsfield ST, Parkhill J, et al. Improved Prediction of Bacterial Genotype-Phenotype Associations Using Interpretable Pangenome-Spanning Regressions. mBio. 2020;11(4). doi:10.1128/mbio.01344-20.

70. Ellington MJ, Ekelund O, Aarestrup FM, Canton R, Doumith M, Giske C, et al. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. Clinical Microbiology and Infection. 2017;23(1):2–22. doi:https://doi.org/10.1016/j.cmi.2016.11.012.

71. Langley G, Besser J, Iwamoto M, Lessa FC, Cronquist A, Skoff TH, et al. Effect of Culture-Independent Diagnostic Tests on Future Emerging Infections Program Surveillance. Emerging Infectious Diseases. 2015;21(9):1582–1588. doi:10.3201/eid2109.150570.

72. Gordon NC, Price JR, Cole K, Everitt R, Morgan M, Finney J, et al. Prediction of Staphylococcus aureus antimicrobial resistance by whole-genome sequencing. Journal of Clinical Microbiology. 2014;52(4):1182–1191. doi:10.1128/JCM.03117-13.

73. Stoesser N, Batty EM, Eyre DW, Morgan M, Wyllie DH, Del Ojo Elias C, et al. Predicting antimicrobial susceptibilities for Escherichia coli and Klebsiella pneumoniae isolates using whole genomic sequence data. Journal of Antimicrobial Chemotherapy. 2013;68(10):2234–2244. doi:10.1093/jac/dkt180.

74. Zankari E, Hasman H, Kaas RS, Seyfarth AM, Agersø Y, Lund O, et al. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. Journal of Antimicrobial Chemotherapy. 2012;68(4):771–777. doi:10.1093/jac/dks496.

75. Köser CU, Ellington MJ, Peacock SJ. Whole-genome sequencing to control antimicrobial resistance. Trends in Genetics. 2014;30(9):401–407. doi:https://doi.org/10.1016/j.tig.2014.07.003.

76. Kim JI, Maguire F, Tsang KK, Gouliouris T, Peacock SJ, McAllister TA, et al. Machine Learning for Antimicrobial Resistance Prediction: Current Practice, Limitations, and Clinical Perspective. Clinical Microbiology Reviews. 2022; p. e0017921–e0017921. doi:10.1128/cmr.00179-21.

77. Paton JH, Reeves DS. Fluoroquinolone Antibiotics. Drugs. 1988;36(2):193–228. doi:10.2165/00003495-198836020-00004.

78. Lindberg F, Normark S. Contribution of Chromosomal Beta-Lactamases to Beta-Lactam Resistance in Enterobacteria. Reviews of Infectious Diseases. 1986;8:S292–S304. doi:10.1093/clinids/8.Supplement_3.S292.

79. Deekshit VK SS. 'To be, or not to be'—The dilemma of 'silent' antimicrobial resistance genes in bacteria. Journal of Applied Bacteriology. 2022;doi:10.1111/jam.15738.

Supporting information

S1 Table. Important predictors in fitted *Gene&Binary&Count* elastic net model by antibiotic. Listed are the important predictors identified by Gene&Binary&Count models for each antibiotic. Predictor importance is estimated by the absolute value of the predictor coefficient in the model. Animal host interaction terms are not shown.

S2 Table. Non-interpretable phenotypes by antibiotic. The number of sample with non-interpretable phenotypes for specific host animal and antibiotics is listed.

S3 Table. Importance of *Gene&Binary&Count* predictors. Predictor importance is calculated as the absolute value of the predictor coefficient in the fitted model. Importance is grouped into main effects, animal, and genetic predictors.

S4 Table. Important predictors for *Gene&Binary&Count* model. Listed are the important predictors identified by Gene&Binary&Count models for each antibiotic are listed. Predictor importance is estimated by the absolute value of the predictor coefficient in the model. Animal host interaction terms are not shown.

S5 Table. *Gene&Binary&Count* model performance. Performance metrics for the *Gene&Binary&Count* elastic net model are presented.