An accurate and interpretable model to predict antimicrobial resistance in One Health settings

Henri C. Chung[1], Christine L. Foxx[2], Jessica A. Hicks[3], Tod 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 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, whole genome sequencing and antibiotic susceptibility test data from *Escherichia coli* were used to identify AMR genotypes and phenotypes for 24 antibiotics. We test the strength of estimated genotype-to-phenotype relationships for 288 AMR genes and associated plasmid replicons with elastic net logistic regression. Model predictors were selected to evaluate different potential modes of AMR genotype translation into resistance phenotypes. We demonstrate that genotypes can be used to accurately predict resistance phenotypes, and there are unique predictors of resistance for different antibiotics within the same class. A model considering the presence of individual genes, total number of genes, a binary indicator of AMR gene group, and host animal most accurately predicted isolate resistance across all tested antibiotics (mean F_1 score = 98%, mean accuracy = 98.2%).

Author's Description

Antimicrobial resistance (AMR) is a major concern in human and animal health. The One Health approach recognizes that human and animal health are linked, and that AMR genes readily transfer between animals and humans in both directions. Transfers can be direct, as in food consumption or contact with farm or companion animals, or indirect, as in farm-based pollution or contact with intermediate hosts. Therefore, it is important to be able to predict AMR in farm and companion animals. Here we constructed an accurate and interpretable machine learning model to predict AMR in *Escherichia coli* (*E. coli*) using isolates from farm and companion animals. The model shows that significant predictors include the presence of individual AMR genes, total number of AMR genes, and host animal species. Moreover, resistance phenotype can be predicted using information from only a subset of possible genetic contributors to resistance, and predictors can vary between antibiotics of the same class, even when we expect predictors to be identical or similar to one another. Understanding these difference enables public health officials, veterinary practitioners and environmental health scientists to make accurate antibiotic specific AMR predictions, and more confidently identify early cases of AMR as they emerge to prevent their spread.

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 their well-being. Multiple organizations including the World Health Organization, the European Commission, and major U.S. agencies not limited to 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 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 issued the first National Action Plan for Combating Antibiotic-Resistant Bacteria (CARB NAP) in 2015, directing all 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]. The 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 methods for sharing and disseminating this information within the veterinary community, and (3) data that would help inform future policies designed to mitigate AMR among animals [9]. We used data from this pilot project to investigate the AMR genotype-to-phenotype relationships.

Genetic models of AMR resistance are an invaluable area of veterinary research. The decreased cost and increased availability of whole genome sequencing (WGS) 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 explosive growth of information have turned to machine learning (ML) models to further enhance AMR surveillance [13]. These models use sequencing information to make accurate predictions of resistance phenotype, alleviating the need for antimicrobial susceptibility testing and potentially identifying new genetic determinants in the process [14].

Despite the impressive number of studies on genotype-to-phenotype relationships and ML prediction for AMR, there is a severe paucity of literature on how these relationships vary depending on the specific antibiotic, animal host, or other clinical factors. While several studies have found 95% to 100% correlation between AMR genotype and phenotype in multiple bacterial pathogens, they also find appreciable differences in the rates of resistance between fluoroquinolones, tetracyclines, aminoglycosides and β-lactams [15–17].

Differences in rates of resistance may be caused by underlying differences in mechanisms that AMR genes use to resist different types of antibiotics, or differences based on the prevalence of their use in veterinary or animal husbandry practice [18, 19]. Bacterial isolates express widely variable rates of resistance depending on the host animal species [20, 21]. At the same time, animal host information has rarely been considered in AMR models, and this variation suggests that modeling AMR genotype-phenotype relationships is more complex than what is currently considered. One 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[22]. 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.

We used WGS data to assess genotype-to-phenotype relationships for multiple antibiotics in *Escherichia coli* isolated from sick animals. Sequenced isolates of *E. coli*, a zoonotic and cross-species pathogen, can serve as indicators of emerging AMR [23, 24]. *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 because it can accumulate AMR genes through multiple mechanisms of horizontal gene transfer [25]. By examining the AMR profiles of *E. coli* isolates interrogated through WGS, we aim to develop an understanding of the different genetic associations which lead to the expression of AMR phenotypes in domesticated animals of interest to the clinical veterinary diagnostic community and public health practitioners.

We used *E. coli* isolates recovered from seven different animal species to determine the utility of WGS-derived AMR genotype data for predicting phenotypic resistance in a clinically relevant context. Isolates came from various species of farm and companion animals as part of the APHIS-NAHLN AMR pilot project from 2018-2022. Our goal was to construct a model that would use AMR gene content combined with host effects 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.

Previous studies using regression methods to predict AMR resistance with WGS data have only focused on model performance. Machine learning methods have been successfully used to predict AMR resistance in pathogens such as non-typhoidal Salmonella, *Mycobacterium tuberculosis*, and *E. coli* [26–33]. While the predictions were accurate, these studies did not attempt to understand the biological factors that could affect predictive accuracy when adopted for large scale use. Model sensitivity to the choice of pathogen, animal host, or differences between antibiotics are often overlooked or simply out of scope for the data available in each individual study. In this study, we show that AMR prediction is highly contextual, and therefore

requires a prediction method that takes animal effects and gene associations into account when predicting AMR for specific antibiotics.

In our models, we included predictors that test ways in which AMR genes could confer resistance. The proposed associations are based on features such as presence/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. These differences suggest future AMR modeling approaches should be more specific, 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

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 collected 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 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 to the genus and species level by each laboratory, typically by using MALDI-TOF methods standard for veterinary

microbiology laboratories [34]. Data submitted for each isolate included the minimum inhibitory concentration (MIC) 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 a duck sample from subsequent analyses as it represented a singleton animal group inappropriate for further, downstream analysis. 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 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* [35]. 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) [36]. Isolates with MIC values at or below the ECOFF value are considered to be members of the wildtype population for that given 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 [37]. 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 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 customized database consisting of RefSeq complete genomes database release ver. 2.09 [38], 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 [39, 40]. Following this validation step, *de novo* assembly of bacterial genomes was performed using SPades (ver. 3.14.0, [41]) and resultant scaffolds were processed using the AMRFinder [42] and ABRicate [43] tool kits against the National Center for Biotechnology Information (NCBI), AMRFinder, ResFinder [44], and PlasmidFinder [45] databases. Resultant gene hits from these tool kits, with minimum inclusion requirements of 70% identity and 95% reference gene coverage, were then collated with the metadata as putative AMR genes found in each *E. coli* isolate. Across all samples, a total of 288 different AMR genes were identified.

Determining animal host differences

We used Fisher's exact test to determine if there were significant differences in the proportion of resistant or non-*wt* phenotypes between animal hosts for each antibiotic in the data set. The Bonferroni-Holm correction for multiple comparisons was applied to p-values, and a corrected p-value threshold of 0.05 was used to determine significance.

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 [46–48]. 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. The shrinking and predictor importance selection make elastic net suited for the task WGS-AMR prediction, where the number of predictors greatly exceeds the number of samples, while retaining a higher level of predictor interpretability compared to other ML methods [49]. 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 constructed a custom database to determine which genes confer resistance to each antibiotic based on manually curated knowledge of AMR genes. In this database, AMR genes can be 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 with resistant (or non-*wt*) and sensitive or *wt* samples fewer than 5. For the remaining 31 antibiotic and breakpoint combinations, we admitted only genes into the relevant set $\mathcal{X}_i$ known to confer resistance to antibiotic i, except in the *Full* model, where all AMR genes, including genes associated with other antibiotics, were considered. We used elastic net with the logistic regression in 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), $\boldsymbol{y}_i^t = (y_{i1}, y_{i2}, \ldots, y_{im_i})$, 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 $\mathbf{\Gamma}_i$ 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.

To explore the genetic association of AMR genes, we considered models built from different sets of genotypic predictors (Table 1), in all cases allowing main and gene×interaction effects for host species. 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 here to represent the different models, with $\boldsymbol{w}_i$ indicating animal, $\boldsymbol{x}_i$ indicating presence/absence of genes in the relevant set $\mathcal{X}_i$ ($\boldsymbol{x}$ for all AMR genes, not just those in the relevant set), $\boldsymbol{v}_i$ indicating 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.

Name	**Model**
Gene	$\boldsymbol{y}_i \sim \boldsymbol{w}_i * \boldsymbol{x}_i$
Gene+Binary	$\boldsymbol{y}_i \sim \boldsymbol{w}_i * (\boldsymbol{x}_i + \boldsymbol{v}_i)$
Gene+Count	$\boldsymbol{y}_i \sim \boldsymbol{w}_i * (\boldsymbol{x}_i + \boldsymbol{z}_i)$
Gene+Binary+Count	$\boldsymbol{y}_i \sim \boldsymbol{w}_i * (\boldsymbol{x}_i + \boldsymbol{v}_i + \boldsymbol{z}_i)$
Full	$\boldsymbol{y}_i \sim \boldsymbol{w}_i * \boldsymbol{x}$
Count	$\boldsymbol{y}_i \sim \boldsymbol{w}_i * \boldsymbol{z}_i$
Binary	$\boldsymbol{y}_i \sim \boldsymbol{w}_i * \boldsymbol{v}_i$

After applying the defined inclusion and exclusion 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 [50, 51]. 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 [52, 53]. Important predictors are the selected predictors found in over 66% or two-thirds 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

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 Positive Predictive Value and True Positive Rate; 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).

There is no commonly 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 [54], which was ignored in our power analysis. We based our analysis on simple logistic regression and a Wald test-derived procedure described in Demidenko [54] and implemented in the R *WebPower* package [54, 55] 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 5 samples. The highest prevalence of resistance were for minocycline (100%; $n = 21$), doxycycline (94.5%; $n = 52$), and ampicillin (92.5%; $n = 571$). The lowest prevalence were 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 agaist amikacin ($n = 175$), azithryomycin ($n = 116$), cephalexin ($n = 152$), or florfenicol ($n = 65$), although all but amikacin have Non-Interpretable MIC values.

Previous studies have shown isolate resistance varies between different animal species [56, 57], 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 1). For streptomycin and neomycin, the rate of resistance was significantly lower in chickens 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 proportion 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 like chickens and turkeys, or companion species like cats and dogs), there are significant differences in rates of resistance observed for a given antibiotic.

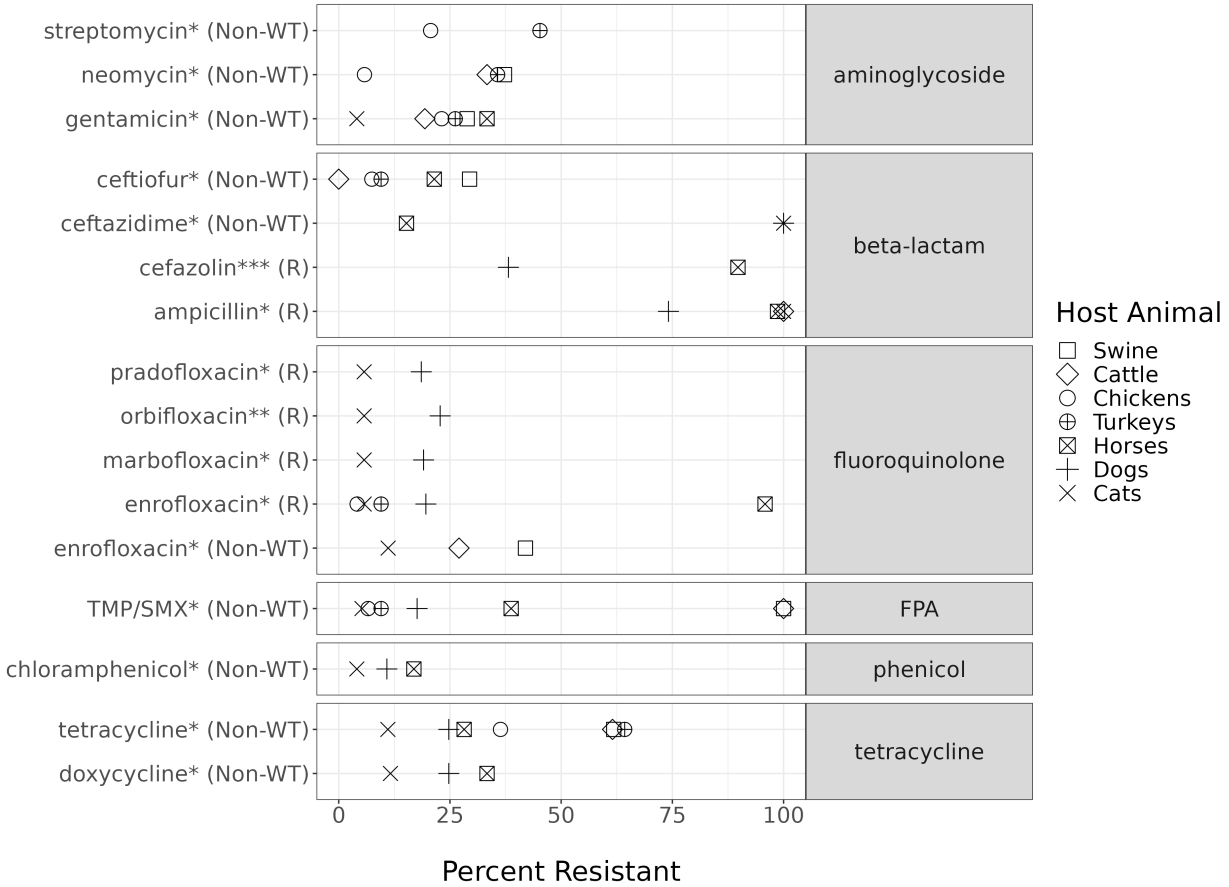

Fig 1. Significant differences in antibiotic resistance between animal hosts. The y-axis labels indicate the antibiotic and breakpoints used to determine resistant phenotype. The x-axis represents the percentage of resistant isolates. Symbols indicate animal host. Each point is the proportion of resistant 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 wanted to examine which trait or combination of traits best predicts AMR resistance. To that end, we fit elastic net models to genotype and phenotype data from multiple *E. coli* isolates. The challenge in these data is the relatively small sample size when considering a large number of potential predictors and their interactions with the animal host. Including all predictors would lead to an over-parameterized model with poor predictive performance. To address this challenge, we use elastic net, a regularized regression method, to avoid overfitting and aid in variable selection. The elastic net combines two regularization methods, Least Absolute Shrinkage and Selection Operator (LASSO) and Ridge regressions, to alleviate the individual drawbacks of each approach when selecting the best predictors. LASSO regression uses a selection process to

identify important predictors from a larger candidate set, but has difficulty differentiating between predictors when they are collinear. Ridge regression handles multicollinearity well, but does not select predictors based on their predictive power. By adding the regression penalties of both methods, an elastic net is able to choose important predictors even when the data is multicollinear. It has been successfully applied to produce interpretable models for AMR in four human pathogens [58].

We fit increasingly complex elastic net models corresponding to types of association for the genetic determinants of AMR (Fig 2). We assume these associations will vary depending on the host and each model includes host animal predictors and interactions as a baseline. Many existing AMR models predict resistance based on the presence of *any* AMR genes relevant to the modeled antibiotic [59–61]. A relevant AMR gene is a gene known to confer resistance to the response antibiotic. We fit an equivalent model, *Binary*, which achieved a mean F_1 of $69.2\% \pm 24.1\%$. We subsequently fit more complex models building upon reasonably presumed genetic associations of resistance. The *Count* model evaluates whether the log odds of resistance linearly increases with the presence of every additional AMR gene, and achieved a mean F_1 of $92.0\% \pm 10.3\%$. The *Gene* model uses the presence of individual AMR genes as separate predictors, allowing different genes to have distinct roles in conferring resistance. This model, despite having more predictors, only achieved a F_1 of $84.1\% \pm 18\%$, lower than either of the previous models. However, the performance of the *Gene* model may have been affected by the power to detect gene-level effects.

Therefore, 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 second best performance with a mean F_1 of $98.2\% \pm 2.27\%$, and was only marginally exceeded by a joint *Gene+Binary+Count* model with a mean F_1 of $98.3\% \pm 2.2\%$. While the performance of the *Gene+Count* and *Gene+Binary+Count* models were nearly equivalent ($98.2\% \pm 2.27\%$ and $98.3\% \pm 2.2\%$, respectively), the addition of the "binary" predictor to the *Gene* model conferred some information on the resistance relationship, but is mostly eclipsed by the more pertinent and predictive information in "count". The improvement in performance confirms, when there is sufficient power, AMR genes do not contribute equally to 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 (F_1 of $88.7\% \pm 14.4\%$).

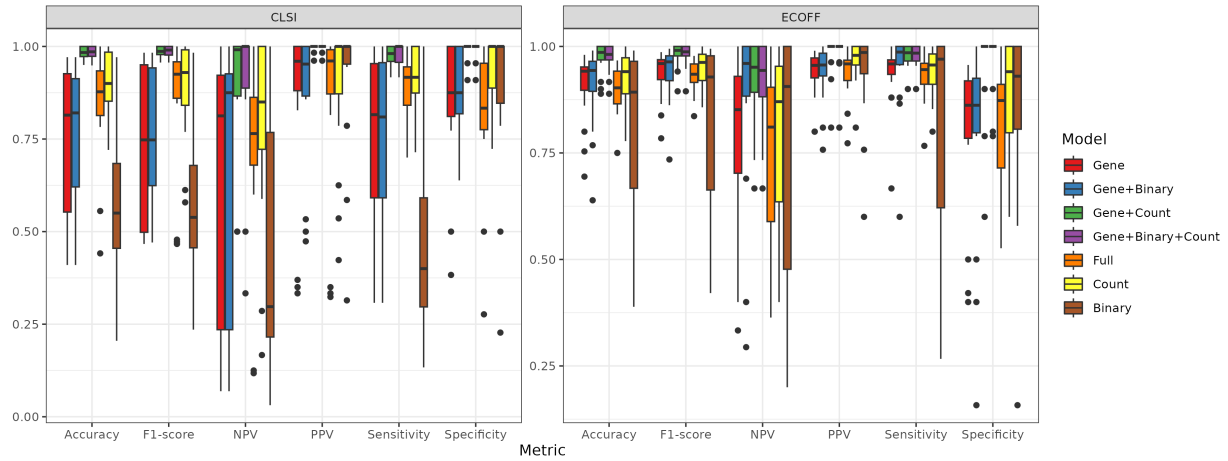

Fig 2. Elastic net model evaluation. We fit seven different elastic net models evaluating the effect of different predictors described in Table 1, including: AMR genes, animal host, presence of $n \geq 1$ AMR gene ("binary"), and total counts of relevant AMR genes (n, "count"). The *Gene+Binary+Count* model best predicts the resistance phenotype across all antibiotics, with the highest median F_1 score and median accuracy (1). 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" or number of genes improved phenotype prediction most compared to additional genes or "binary" classification. The *Count* model achieved performance exceeding more complex models while only using a single predictor. The importance of "count" suggests the number of resistance genes is more important for predicting resistance than gene identity, and individual genes alone cannot be used predict resistance. 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 "count" 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. 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 F_1 of $98.0\% \pm 2.4\%$. Despite our initial assumptions, the inclusion of animal predictors did not have a significant effect on resistance prediction. 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 collinearity between different AMR genes and animal hosts would allow genetic predictors to effectively substitute for animal effects [58]. A no-animal *Count* model, which masked the identity of individual genes

and did not use any animal information, achieved a mean F_1 score of $90.5\% \pm 14.7\%$.

Analysis of individual predictors

We studied the individual predictors used in the *Gene+Binary+Count* model to predict antibiotic resistance. For the *Gene+Binary+Count* models, we calculated the important predictors for each antibiotic and compared them to important predictors for other antibiotics in the same class. Although the inclusion of "binary" classifications, animal predictors, and their interactions did little to improve model performance, their importance suggests they bear some effect on the antibiotic phenotype. The shrinkage effect of the elastic net would minimize the coefficients of potentially biologically meaningful predictors if they are collinear with another feature. Due to this possibility, and the flexibility of elastic net to additional predictors, we chose to retain them in our final model. 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 inevitably under-powered for many possible predictors. Thus, we computed for each predictor a power to detect with a simplified model (see Methods). Across 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%).

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. We observe 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 (66.6%) of resistance models did not use any. While models used animal information, the total importance of animal predictors in the *Gene+Binary+Count* model was relatively low (Table S3).

Unique predictors for resistance within an antibiotic class.

We compared the number of important predictors for each model to identify differences in model complexity between antibiotics. 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 2). 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 to the varying complexity of β-lactam models, the models for fluoroquinolones were much simpler. 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 ($n = 11$) available for the model to select from. However, the models for resistant and non-*wt* enrofloxacin have a comparable number of candidate predictors to larger models 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, 22/30 (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 2. 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
		gentamicin		36	41	110	146	146
		neomycin		15	24	56.5	64	64
		streptomycin		28	40	94.9	47	47
		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
		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 against antibiotics within the same class could indicate that differences do exist in the genetic associations for resistance, despite the antibiotics sharing a similar chemical structure. 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 WGS has been proposed as a major tool for monitoring antibiotic resistance [62–64]. Multiple studies have used WGS to predict resistance phenotypes in pathogens such as *E. coli*, *Klebsiella pneumoniae*, *Salmonella enterica* serovar Typhimurium, and *Staphylococcus aureus* [65–67]. These studies report high concordance between a pathogen‘s AMR profile and predicted resistance phenotype, and affirm the potential of sequencing methods to improve AMR surveillance. However, we found these AMR genotype-phenotype relationships to vary substantially depending on the animal host and antibiotic, indicating a degree of AMR complexity that has not yet been discussed. These results are supported by earlier studies which show the varying rates of resistance for antibiotics across animal hosts [20, 21]. While the potential utility and use of WGS 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.

Our results show that the relationship between AMR genotype and phenotype varies depending on the antibiotic and animal host. We compared multiple models to evaluate the effect of different predictors when modeling AMR, and outline the complexity of genotype-to-phenotype resistance to a degree not shown in previous AMR ML projects [68–70]. Our rule-based or *Binary* model, using only a binary indicator of one or more relevant AMR genes, was the worst performing model an mean F_1 score of 69.2%. We subsequently introduced two additional predictors: “binary“ which represents the effect of genes within a group, and “count“ which represents the summative count of total genes found in the *E. coli* isolate. The “count” predictor assumes an AMR model where the probability of resistance linearly increases with the presence of each additional relevant gene. Inclusion of the “count” predictor improved model performance across all evaluation metrics (accuracy, F_1 score, NPV, PPV, sensitivity, and specificity).

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 powerful predictor of resistance than the presence or absence of the genes themselves. 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 [71]. For these

antibiotics, the number of genes and, in turn, the number of synthesized proteins, may be a more accurate predictor of resistance than individual genes themselves. Another possibility is that the model was over-parameterized, and a more parsimonious model with a single predictor could achieve better performance. Furthermore, additional samples would be necessary to properly validate potential biological mechanisms of AMR and to address potential problems of model complexity. While the number or count of relevant AMR genes is a powerful predictor of resistance that exceeds the rule-based model 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 [72]. The model for amoxicillin-clavulanic acid resistance, 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.

Similarly, fluoroquinolone class predictions relied entirely on a single "count" variable. Point mutations are major contributors to fluoroquinolone resistance-determining regions [73]. The importance of "count" for fluoroquinolone prediction could represent the likelihood of resistance increasing with the number of point mutations within these regions. The difference in model complexity for each antibiotic type suggests there are more complex, antibiotic-specific AMR genotype-to-phenotype relationships than have been previously explored [15–17]. 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 observe a difference in the selection for animal predictors in non-*wt* and resistant phenotypes. Animal predictors and interactions are more commonly found in resistant models and are expected, as the ECOFF values used to determine non-*wt* phenotypes are not animal-specific nor purpose-built for intended use in clinical applications. 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 F_1 for both models were very high (98.8). 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 [74]. We note that the models placed different weights on predictors which represent alleles or subclasses of the same gene, suggesting unique allelic differences in resistance transfer that have previously not been observed to the knowledge of the authors. An essential limitation for further inference into these allelic differences is the power behind identified predictors. Theoretical power calculations for elastic net models have yet to be defined, but calculations based on traditional logistic regression suggest that this observational study is under-powered for a majority of predictors at effect sizes less than 2.3 or a ten-fold increases in the odds of resistance. Insufficient power is expected in observational studies, where both antibiotic states and genes can be severely imbalanced. Important predictors of AMR would best be assessed in an experiment explicitly designed to detect fthe 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. Following the emergence of WGS data, there has been advocacy for the use of ML models to predict AMR phenotypes [13, 75]. While our results agree with previous research demonstrating the potential of ML for AMR prediction, the findings are limited by the inherent caveats of ML methods: specifically, the dependence on training samples and interpretation of input

features.

Conclusions

Our study illustrates the genetic complexity of AMR that is often ignored in ML approaches. 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 potentially, allelic differences in 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 resistance prediction, 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 the effect size of individual genes can more accurately be measured. A more accurate measure of effect size can then be used to appropriately weight predictors and improve models for AMR. While there is considerable evidence that WGS data will improve AMR prevention strategies, our results emphasize a need to better understand the relationship between WGS and antibiotic resistance phenotypes. With further improvement, predictive models built on these relationships can prove to be an invaluable for early diagnosis and prevention of emerging AMR.

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

Funding

This research was supported in part by the USDA with internal funds as part of routine work, by the USDA National Institute of Food and Agriculture (Hatch Project IOW03717), and by an appointment to the Animal and Plant Health Inspection Service (APHIS) Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the USDA. ORISE is managed by the Office of Research in Associated Universities (ORAU) under DOE contract number DE-SC0014664. IF gratefully acknowledges funding from the Translational Artificial Intelligence Center (TrAC) from Iowa State University.

Conflicts of interest

All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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] World Health Organization. Global action plan on antimicrobial resistance; World Health Organization: Geneva, Switzerland. Technical report, World Health Organization, Geneva, Switzerland, 2016.

[2] Smith Elta, Lichten Catherine A., Taylor Jirka, MacLure Calum, Lepetit Louise, Harte Emma, Martin Adam, Ghiga Ioana, Pitchforth Emma, Sussex Jon, Dujso (RAND Europe) Elma, and Littmann Jasper. Evaluation of the action plan against the rising threats from antimicrobial resistance. Technical report, European Comission, Brussels, Belgium, 2016.

[3] United States Center for Disease Control. Antibiotic resistance threats in the United States, 2019. Technical report, 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. URL `https://obamawhitehouse.archives.gov/sites/default/files/docs/national_action_plan_for_combating_antibotic-resistant_bacteria.pdf`.

[6] Allison White and James M Hughes. Critical importance of a One Health approach to antimicrobial resistance. EcoHealth, 16(3):404–409, 2019. ISSN 1612-9202. doi: 10.1007/s10393-019-01415-5.

[7] Beth E. Karp, Heather Tate, Jodie R. Plumblee, Uday Dessai, Jean M. Whichard, Eileen L. Thacker, Kis Robertson Hale, Wanda Wilson, Cindy R. Friedman, Patricia M. Griffin, and Patrick F. McDermott. National antimicrobial resistance monitoring system: Two decades of advancing public health through integrated surveillance of antimicrobial resistance. Foodborne Pathogens and Disease, 14(10):545–557, 2017. doi: 10.1089/fpd.2017.2283. PMID: 28792800.

[8] U.S. Food and Drug Administration. Timeline of FDA action on antimicrobial resistance, 2022. URL `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. Technical report, United States Food and Drug Administration, 2018.

[10] Manish Boolchandani, Alaric W. D'Souza, and Gautam Dantas. Sequencing-based methods and resources to study antimicrobial resistance. Nature Reviews Genetics, 20:356–370, 2019. doi: 10.1038/s41576-019-0108-4.

[11] Andrew G McArthur and Kara K Tsang. Antimicrobial resistance surveillance in the genomic age. Ann. N. Y. Acad. Sci., 1388(1):78–91, January 2017.

[12] Roby P Bhattacharyya, Nirmalya Bandyopadhyay, Peijun Ma, Sophie S Son, Jamin Liu, Lorrie L He, Lidan Wu, Rustem Khafizov, Rich Boykin, Gustavo C Cerqueira, Alejandro Pironti, Robert F Rudy, Milesh M Patel, Rui Yang, Jennifer Skerry, Elizabeth Nazarian, Kimberly A Musser, Jill Taylor, Virginia M Pierce, Ashlee M Earl, Lisa A Cosimi, Noam Shoresh, Joseph Beechem, Jonathan Livny, and Deborah T Hung. Simultaneous detection of genotype and phenotype enables rapid and accurate antibiotic susceptibility determination. Nat. Med., 25(12):1858–1864, December 2019.

[13] Melis N. Anahtar, Jason H. Yang, and Sanjat Kanjilal. Applications of machine learning to the problem of antimicrobial resistance: an emerging model for translational research. Journal of clinical microbiology, 59(7):e0126020–e0126020, 2021. ISSN 0095-1137. doi: https://doi.org/10.1128/JCM.01260-20.

[14] Patrick F McDermott and James J Davis. Predicting antimicrobial susceptibility from the bacterial genome: A new paradigm for one health resistance monitoring. J. Vet. Pharmacol. Ther., 44(2): 223–237, March 2021.

[15] Gregory H. Tyson, Patrick F. McDermott, Cong Li, Yuansha Chen, Daniel A. Tadesse, Sampa Mukherjee, Sonya Bodeis-Jones, Claudine Kabera, Stuart A. Gaines, Guy H. Loneragan, Tom S. Edrington, Mary Torrence, Dayna M. Harhay, and Shaohua Zhao. WGS accurately predicts antimicrobial resistance in Escherichia coli. Journal of Antimicrobial Chemotherapy, 70(10):2763–2769, 07 2015. ISSN 0305-7453. doi: 10.1093/jac/dkv186. URL `https://doi.org/10.1093/jac/dkv186`.

[16] Patrick F McDermott, Gregory H Tyson, Claudine Kabera, Yuansha Chen, Cong Li, Jason P Folster, Sherry L Ayers, Claudia Lam, Heather P Tate, and Shaohua Zhao. Whole-genome sequencing for detecting antimicrobial resistance in nontyphoidal Salmonella. Antimicrobial agents and chemotherapy, 60(9):5515–5520, 2016. ISSN 0066-4804. doi: https://doi.org/10.1128/AAC.01030-16.

[17] S Zhao, G H Tyson, Y Chen, C Li, S Mukherjee, S Young, C Lam, J P Folster, J M Whichard, and P F McDermott. Whole-genome sequencing analysis accurately predicts antimicrobial resistance phenotypes

in Campylobacter spp. Applied and environmental microbiology, 82(2):459–466, 2016. ISSN 0099-2240. doi: 10.1128/AEM.02873-15.

[18] Institute of Medicine and National Research Council, editor. The Use of Drugs in Food Animals. National Academies Press, January 1999. doi: 10.17226/5137. URL https://doi.org/10.17226/5137.

[19] Nataliya Roth, Annemarie Käsbohrer, Sigrid Mayrhofer, Ulrike Zitz, Charles Hofacre, and Konrad J Domig. The application of antibiotics in broiler production and the resulting antibiotic resistance in Escherichia coli: A global overview. Poultry Science, 98(4):1791–1804, 2019. ISSN 0032-5791. doi: https://doi.org/10.3382/ps/pey539. URL https://www.sciencedirect.com/science/article/pii/S0032579119301099.

[20] Lourdes Garcia-Migura, Rene S. Hendriksen, Lorenzo Fraile, and Frank M. Aarestrup. Antimicrobial resistance of zoonotic and commensal bacteria in Europe: The missing link between consumption and resistance in veterinary medicine. Veterinary Microbiology, 170(1):1–9, 2014. ISSN 0378-1135. doi: https://doi.org/10.1016/j.vetmic.2014.01.013. URL https://www.sciencedirect.com/science/article/pii/S0378113514000480.

[21] Ghada Abdelwahab, Hassan Zackaria Ishag, Zulaikha Hammadi, Saeed Yammahi, Mohd Yusof, Muna Yassi, Shaikha neyadi, Asma Mansoori, Fawzia Hamadi, Ibtesam Hamadi, Mohamed Hosani, and Salama Muhairi. Antibiotics resistance in Escherichia coli isolated from from livestock in the Emirate of Abu Dhabi, UAE, 2014–2019. International Journal of Microbiology, 2022, 04 2022. doi: 10.1155/2022/3411560.

[22] Allison L Hicks, Nicole Wheeler, Leonor Sánchez-Busó, Jennifer L Rakeman, Simon R Harris, and Yonatan H Grad. Evaluation of parameters affecting performance and reliability of machine learning-based antibiotic susceptibility testing from whole genome sequencing data. PLoS Comput. Biol., 15(9):e1007349, September 2019.

[23] Kamelia M Osman, Anthony D Kappell, Mohamed Elhadidy, Fatma ElMougy, Wafaa A Abd El-Ghany, Ahmed Orabi, Aymen S Mubarak, Turki M Dawoud, Hassan A Hemeg, Ihab M I Moussa, Ashgan M Hessain, and Hend M Y Yousef. Poultry hatcheries as potential reservoirs for antimicrobial-resistant escherichia coli: A risk to public health and food safety. Sci. Rep., 8(1):5859, April 2018.

[24] Elena A Oniciuc, Eleni Likotrafiti, Adrián Alvarez-Molina, Miguel Prieto, Jesús A Santos, and Avelino Alvarez-Ordóñez. 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), 9(5), May 2018.

[25] Laurent Poirel, Jean-Yves Madec, Agnese Lupo, Anne-Kathrin Schink, Nicolas Kieffer, Patrice Nordmann, and Stefan Schwarz. Antimicrobial resistance in Escherichia coli. Microbiology spectrum, 6 (4), 2018. ISSN 2165-0497. doi: 10.1128/microbiolspec.ARBA-0026-2017.

[26] Katherine E. Niehaus, Timothy M. Walker, Derrick W. Crook, Tim E. A. Peto, and David A. Clifton. Machine learning for the prediction of antibacterial susceptibility in mycobacterium tuberculosis. In IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI). IEEE, June 2014. doi: 10.1109/bhi.2014.6864440. URL `https://doi.org/10.1109/bhi.2014.6864440`.

[27] Mitchell W. Pesesky, Tahir Hussain, Meghan Wallace, Sanket Patel, Saadia Andleeb, Carey-Ann D. Burnham, and Gautam Dantas. 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, 7, November 2016. doi: 10.3389/fmicb.2016.01887. URL `https://doi.org/10.3389/fmicb.2016.01887`.

[28] David W. Eyre, Dilrini De Silva, Kevin Cole, Joanna Peters, Michelle J. Cole, Yonatan H. Grad, Walter Demczuk, Irene Martin, Michael R. Mulvey, Derrick W. Crook, A. Sarah Walker, Tim E. A. Peto, and John Paul. WGS to predict antibiotic MICs for neisseria gonorrhoeae. Journal of Antimicrobial Chemotherapy, 72(7):1937–1947, March 2017. doi: 10.1093/jac/dkx067. URL `https://doi.org/10.1093/jac/dkx067`.

[29] Yang Yang, Katherine E Niehaus, Timothy M Walker, Zamin Iqbal, A Sarah Walker, Daniel J Wilson, Tim E A Peto, Derrick W Crook, E Grace Smith, Tingting Zhu, and David A Clifton. Machine learning for classifying tuberculosis drug-resistance from DNA sequencing data. Bioinformatics, 34(10): 1666–1671, 12 2017. ISSN 1367-4803. doi: 10.1093/bioinformatics/btx801. URL `https://doi.org/10.1093/bioinformatics/btx801`.

[30] Yang Yang, Timothy M Walker, A Sarah Walker, Daniel J Wilson, Timothy E A Peto, Derrick W Crook, Farah Shamout, Irena Arandjelovic, Iñaki Comas, Maha R Farhat, Qian Gao, Vitali Sintchenko, Dick van Soolingen, Sarah Hoosdally, Ana L Gibertoni Cruz, Joshua Carter, Clara Grazian, Sarah G Earle, Samaneh Kouchaki, Yang Yang, Timothy M Walker, Philip W Fowler, David A Clifton, Zamin Iqbal, Martin Hunt, E Grace Smith, Priti Rathod, Lisa Jarrett, Daniela Matias, Daniela M Cirillo, Emanuele Borroni, Simone Battaglia, Arash Ghodousi, Andrea Spitaleri, Andrea Cabibbe, Sabira

Tahseen, Kayzad Nilgiriwala, Sanchi Shah, Camilla Rodrigues, Priti Kambli, Utkarsha Surve, Rukhsar Khot, Stefan Niemann, Thomas Kohl, Matthias Merker, Harald Hoffmann, Nikolay Molodtsov, Sara Plesnik, Nazir Ismail, Guy Thwaites, Thuong Nguyen Thuy Thuong, Nhung Hoang Ngoc, Vijay Srinivasan, David Moore, David Jorge Coronel, Walter Solano, George F Gao, Guangxue He, Yanlin Zhao, Aijing Ma, Chunfa Liu, Baoli Zhu, Ian Laurenson, Pauline Claxton, Anastasia Koch, Robert Wilkinson, Ajit Lalvani, James Posey, James Jennifer Gardy, Jim Werngren, Nicholas Paton, Ruwen Jou, Mei-Hua Wu, Wan-Hsuan Lin, Lucilaine Ferrazoli, Rosangela Siqueira de Oliveira, São Paulo, Tingting Zhu, and David A Clifton and. DeepAMR for predicting co-occurrent resistance of mycobacterium tuberculosis. Bioinformatics, 35(18):3240–3249, January 2019. doi: 10.1093/bioinformatics/btz067. URL https://doi.org/10.1093/bioinformatics/btz067.

[31] Marcus Nguyen, S. Wesley Long, Patrick F. McDermott, Randall J. Olsen, Robert Olson, Rick L. Stevens, Gregory H. Tyson, Shaohua Zhao, and James J. Davis. Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal salmonella. Journal of Clinical Microbiology, 57(2), February 2019. doi: 10.1128/jcm.01260-18. URL https://doi.org/10.1128/jcm.01260-18.

[32] Finlay Maguire, Muhammad Attiq Rehman, Catherine Carrillo, Moussa S. Diarra, and Robert G. Beiko. Identification of primary antimicrobial resistance drivers in agricultural nontyphoidal salmonella enterica serovars by using machine learning. mSystems, 4(4), August 2019. doi: 10.1128/msystems.00211-19. URL https://doi.org/10.1128/msystems.00211-19.

[33] Nicole Pearcy, Yue Hu, Michelle Baker, Alexandre Maciel-Guerra, Ning Xue, Wei Wang, Jasmeet Kaler, Zixin Peng, Fengqin Li, and Tania Dottorini. Genome-scale metabolic models and machine learning reveal genetic determinants of antibiotic resistance in escherichia coli and unravel the underlying metabolic adaptation mechanisms. mSystems, 6(4), August 2021. doi: 10.1128/msystems.00913-20. URL https://doi.org/10.1128/msystems.00913-20.

[34] Neelja Singhal, Manish Kumar, Pawan K. Kanaujia, and Jugsharan S. Virdi. MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. Frontiers in Microbiology, 6, August 2015. doi: 10.3389/fmicb.2015.00791. URL https://doi.org/10.3389/fmicb.2015.00791.

[35] Clinical and Laboratory Standards Institute. Performance standards for antimicrobial disk and dilution

susceptibility tests for bacteria isolated from animals. CLSI supplement, VET08 v. 38, no. 14. Clinical and Laboratory Standards Institute, Wayne, PA, 4th edition. edition, 2018. ISBN 9781684400102.

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

[37] G. Kahlmeter. European harmonization of MIC breakpoints for antimicrobial susceptibility testing of bacteria. Journal of Antimicrobial Chemotherapy, 52(2):145–148, July 2003. doi: 10.1093/jac/dkg312. URL https://doi.org/10.1093/jac/dkg312.

[38] Tatiana Tatusova, Michael DiCuccio, Azat Badretdin, Vyacheslav Chetvernin, Eric P. Nawrocki, Leonid Zaslavsky, Alexandre Lomsadze, Kim D. Pruitt, Mark Borodovsky, and James Ostell. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Research, 44(14):6614–6624, June 2016. doi: 10.1093/nar/gkw569. URL https://doi.org/10.1093/nar/gkw569.

[39] Jennifer Lu, Florian P Breitwieser, Peter Thielen, and Steven L Salzberg. Bracken: estimating species abundance in metagenomics data. PeerJ. Computer science, 3:e104–, 2017. ISSN 2376-5992. doi: 10.7717/peerj-cs.104.

[40] Jennifer Lu, Natalia Rincon, Derrick E. Wood, Florian P. Breitwieser, Christopher Pockrandt, Ben Langmead, Steven L. Salzberg, and Martin Steinegger. Metagenome analysis using the Kraken software suite. Nature Protocols, 17(12):2815–2839, September 2022. doi: 10.1038/s41596-022-00738-y. URL https://doi.org/10.1038/s41596-022-00738-y.

[41] Sergey Nurk, Anton Bankevich, Dmitry Antipov, Alexey A. Gurevich, Anton Korobeynikov, Alla Lapidus, Andrey D. Prjibelski, Alexey Pyshkin, Alexander Sirotkin, Yakov Sirotkin, Ramunas Stepanauskas, Scott R. Clingenpeel, Tanja Woyke, Jeffrey S. Mclean, Roger Lasken, Glenn Tesler, Max A. Alekseyev, and Pavel A. Pevzner. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. Journal of Computational Biology, 20(10):714–737, 2013. doi: 10.1089/cmb.2013.0084. URL https://doi.org/10.1089/cmb.2013.0084. PMID: 24093227.

[42] Michael Feldgarden, Vyacheslav Brover, Daniel H Haft, Arjun B Prasad, Douglas J Slotta, Igor Tolstoy, Gregory H Tyson, Shaohua Zhao, Chih-Hao Hsu, Patrick F McDermott, 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, 63(11):e00483–19, 2019. ISSN 0032-5791. doi: https://doi.org/10.1093/ps/84.4.634. URL https://pubmed.ncbi.nlm.nih.gov/31427293/.

[43] Torsten Seemann. Abricate. https://github.com/tseemann/abricate, 2014.

[44] Ea Zankari, Henrik Hasman, Salvatore Cosentino, Martin Vestergaard, Simon Rasmussen, Ole Lund, Frank M Aarestrup, and Mette Voldby Larsen. Identification of acquired antimicrobial resistance genes. Journal of antimicrobial chemotherapy, 67(11):2640–2644, 2012. ISSN 0305-7453. doi: 10.1093/jac/dks261.

[45] Alessandra Carattoli, Ea Zankari, Aurora García-Fernández, Mette Voldby Larsen, Ole Lund, Laura Villa, Frank Møller Aarestrup, and Henrik Hasman. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrobial agents and chemotherapy, 58(7): 3895–3903, 2014. ISSN 0032-5791. doi: https://doi.org/10.1128/aac.02412-14. URL https://pubmed.ncbi.nlm.nih.gov/24777092/.

[46] Robert Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58(1):267–288, 1996. ISSN 00359246. URL http://www.jstor.org/stable/2346178.

[47] Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 42(1):80–86, 2000. ISSN 00401706. URL http://www.jstor.org/stable/1271436.

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

[49] Jee In Kim, Finlay Maguire, Kara K. Tsang, Theodore Gouliouris, Sharon J. Peacock, Tim A. McAllister, Andrew G. McArthur, and Robert G. Beiko. Machine learning for antimicrobial resistance prediction: Current practice, limitations, and clinical perspective. Clinical Microbiology Reviews, 35(3), September 2022. doi: 10.1128/cmr.00179-21. URL https://doi.org/10.1128/cmr.00179-21.

[50] Chenlei Leng, Yi Lin, and Grace Wahba. A note on the lasso and related procedures in model selection. Statistica Sinica, pages 1273–1284, 2006.

[51] Christophe Ambroise and Geoffrey J. McLachlan. Selection bias in gene extraction on the basis of microarray gene-expression data. Proceedings of the National Academy of Sciences, 99(10):6562–6566, April 2002. doi: 10.1073/pnas.102102699. URL https://doi.org/10.1073/pnas.102102699.

[52] Md Hasinur Rahaman Khan, Anamika Bhadra, and Tamanna Howlader. Stability selection for lasso, ridge and elastic net implemented with AFT models. Statistical Applications in Genetics and Molecular Biology, 18(5), October 2019. doi: 10.1515/sagmb-2017-0001. URL https://doi.org/10.1515/sagmb-2017-0001.

[53] Nicolai Meinshausen and Peter Bühlmann. Stability selection. J. R. Stat. Soc. Series B Stat. Methodol., 72(4):417–473, July 2010.

[54] Eugene Demidenko. Sample size determination for logistic regression revisited. Statistics in Medicine, 26(18):3385–3397, December 2007. doi: 10.1002/sim.2771. URL https://doi.org/10.1002/sim.2771.

[55] Zhiyong Zhang and Ke-Hai Yuan. Practical statistical power analysis using webpower and R. ISDSA Press, 2018.

[56] Roland Lanz, Peter Kuhnert, and Patrick Boerlin. Antimicrobial resistance and resistance gene determinants in clinical Escherichia coli from different animal species in Switzerland. Veterinary Microbiology, 91(1):73–84, 2003. ISSN 0378-1135. doi: https://doi.org/10.1016/S0378-1135(02)00263-8. URL https://www.sciencedirect.com/science/article/pii/S0378113502002638.

[57] Clémence Bourély, Thomas Coeffic, Jocelyne Caillon, Sonia Thibaut, Géraldine Cazeau, Eric Jouy, Nathalie Jarrige, Claire Chauvin, Jean-Yves Madec, Marisa Haenni, Agnès Leblond, and Emilie Gay. Trends in antimicrobial resistance among Escherichia coli from defined infections in humans and animals. Journal of Antimicrobial Chemotherapy, 75(6):1525–1529, 02 2020. ISSN 0305-7453. doi: 10.1093/jac/dkaa022. URL https://doi.org/10.1093/jac/dkaa022.

[58] John A. Lees, T. Tien Mai, Marco Galardini, Nicole E. Wheeler, Samuel T. Horsfield, Julian Parkhill, and Jukka Corander. Improved prediction of bacterial genotype-phenotype associations using interpretable pangenome-spanning regressions. mBio, 11(4), August 2020. doi: 10.1128/mbio.01344-20. URL https://doi.org/10.1128/mbio.01344-20.

[59] Kara K. Tsang, Finlay Maguire, Haley L. Zubyk, Sommer Chou, Arman Edalatmand, Gerard D. Wright, Robert G. Beiko, and Andrew G.YR 2021 McArthur. Identifying novel β-lactamase substrate

activity through in silico prediction of antimicrobial resistance. Microbial Genomics, 7(1):000500, 2021. doi: 10.1099/mgen.0.000500.

[60] Gregory H. Tyson, Olgica Ceric, Jake Guag, Sarah Nemser, Stacey Borenstein, Durda Slavic, Sarah Lippert, Rebecca McDowell, Aparna Krishnamurthy, Shannon Korosec, Cheryl Friday, Neil Pople, Matthew E. Saab, Julie-Hélène Fairbrother, Isabelle Janelle, Deanna McMillan, Yugendar R. Bommineni, David Simon, Shipra Mohan, Susan Sanchez, Ashley Phillips, Paula Bartlett, Hemant Naikare, Cynthia Watson, Orhan Sahin, Chloe Stinman, Leyi Wang, Carol Maddox, Vanessa DeShambo, G. Kenitra Hendrix, Debra Lubelski, Amy Burklund, Brian Lubbers, Debbie Reed, Tracie Jenkins, Erdal Erol, Mukeshbhai Patel, Stephan Locke, Jordan Fortner, Laura Peak, Udeni Balasuriya, Rinosh Mani, Niesa Kettler, Karen Olsen, Shuping Zhang, Zhenyu Shen, Martha Pulido Landinez, Jay Kay Thornton, Anil Thachil, Melissa Byrd, Megan Jacob, Darlene Krogh, Brett Webb, Lynn Schaan, Amar Patil, Sarmila Dasgupta, Shannon Mann, Laura B. Goodman, Rebecca June Franklin-Guild, Renee R. Anderson, Patrick K. Mitchell, Brittany D. Cronk, Missy Aprea, Jing Cui, Dominika Jurkovic, Melanie Prarat, Yan Zhang, Katherine Shiplett, Dubraska Diaz Campos, Joany Van Balen Rubio, Akhilesh Ramanchandran, Scott Talent, Deepanker Tewari, Nagaraja Thirumalapura, Donna Kelly, Denise Barnhart, Lacey Hall, Shelley Rankin, Jaclyn Dietrich, Stephen Cole, Joy Scaria, Linto Antony, Sara D. Lawhon, Jing Wu, Christine McCoy, Kelly Dietz, Rebecca Wolking, Trevor Alexander, Claire Burbick, and Renate Reimschuessel. Genomics accurately predicts antimicrobial resistance in *Staphylococcus pseudintermedius* collected as part of Vet-LIRN resistance monitoring. Veterinary Microbiology, 254: 109006, Mar 2021. doi: 10.1016/j.vetmic.2021.109006.

[61] Walter Demczuk, Irene Martin, Averil Griffith, Brigitte Lefebvre, Allison McGeer, Gregory J. Tyrrell, George G. Zhanel, Julianne V. Kus, Linda Hoang, Jessica Minion, Paul Van Caeseele, Rita Raafat Gad, David Haldane, George Zahariadis, Kristen Mead, Laura Steven, Lori Strudwick, and Michael R. Mulvey. Linear regression equations to predict β-lactam, macrolide, lincosamide, and fluoroquinolone MICs from molecular antimicrobial resistance determinants in *Streptococcus pneumoniae*. Antimicrobial Agents and Chemotherapy, 66(1):e01370–21, Jan 2022. doi: 10.1128/AAC.01370-21.

[62] M.J. Ellington, O. Ekelund, F.M. Aarestrup, R. Canton, M. Doumith, C. Giske, H. Grundman, H. Hasman, M.T.G. Holden, K.L. Hopkins, J. Iredell, G. Kahlmeter, C.U. Köser, A. MacGowan, D. Mevius, M. Mulvey, T. Naas, T. Peto, J.-M. Rolain, Ø. Samuelsen, and N. Woodford. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST subcommittee. Clinical Microbiology and Infection, 23(1):2–22, 2017. ISSN 1198-743X. doi:

https://doi.org/10.1016/j.cmi.2016.11.012. URL https://www.sciencedirect.com/science/article/pii/S1198743X16305687.

[63] Gayle Langley, John Besser, Martha Iwamoto, Fernanda C Lessa, Alicia Cronquist, Tami H Skoff, Sandra Chaves, Dave Boxrud, Robert W Pinner, and Lee H Harrison. Effect of culture-independent diagnostic tests on future emerging infections program surveillance. Emerging infectious diseases, 21(9): 1582–1588, 2015. ISSN 1080-6040. doi: 10.3201/eid2109.150570.

[64] N C Gordon, J R Price, K Cole, R Everitt, M Morgan, J Finney, A M Kearns, B Pichon, B Young, D J Wilson, M J Llewelyn, J Paul, T E A Peto, D W Crook, A S Walker, and T Golubchik. Prediction of Staphylococcus aureus antimicrobial resistance by whole-genome sequencing. Journal of clinical microbiology, 52(4):1182–1191, 2014. ISSN 0095-1137. doi: 10.1128/JCM.03117-13.

[65] N Stoesser, E M Batty, D W Eyre, M Morgan, D H Wyllie, C Del Ojo Elias, J R Johnson, A S Walker, T E A Peto, and D W Crook. Predicting antimicrobial susceptibilities for Escherichia coli and Klebsiella pneumoniae isolates using whole genomic sequence data. Journal of antimicrobial chemotherapy, 68 (10):2234–2244, 2013. ISSN 0305-7453. doi: 10.1093/jac/dkt180.

[66] Ea Zankari, Henrik Hasman, Rolf Sommer Kaas, Anne Mette Seyfarth, Yvonne Agersø, Ole Lund, Mette Voldby Larsen, and Frank M. Aarestrup. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. Journal of Antimicrobial Chemotherapy, 68(4):771–777, 12 2012. ISSN 0305-7453. doi: 10.1093/jac/dks496. URL https://doi.org/10.1093/jac/dks496.

[67] Claudio U. Köser, Matthew J. Ellington, and Sharon J. Peacock. Whole-genome sequencing to control antimicrobial resistance. Trends in Genetics, 30(9):401–407, 2014. ISSN 0168-9525. doi: https://doi.org/10.1016/j.tig.2014.07.003. URL https://www.sciencedirect.com/science/article/pii/S0168952514001140.

[68] Zhichang Liu, Dun Deng, Huijie Lu, Jian Sun, Luchao Lv, Shuhong Li, Guanghui Peng, Xianyong Ma, Jiazhou Li, Zhenming Li, Ting Rong, and Gang Wang. Evaluation of machine learning models for predicting antimicrobial resistance of Actinobacillus pleuropneumoniae from whole genome sequences. Frontiers in Microbiology, 11, 2020. ISSN 1664-302X. doi: 10.3389/fmicb.2020.00048. URL https://www.frontiersin.org/articles/10.3389/fmicb.2020.00048.

[69] Danesh Moradigaravand, Martin Palm, Anne Farewell, Ville Mustonen, Jonas Warringer, and Leopold Parts. Prediction of antibiotic resistance in Escherichia coli from large-scale pan-genome data. PLOS

Computational Biology, 14(12):1–17, 12 2018. doi: 10.1371/journal.pcbi.1006258. URL https://doi.org/10.1371/journal.pcbi.1006258.

[70] Hsuan-Lin Her and Yu-Wei Wu. A pan-genome-based machine learning approach for predicting antimicrobial resistance activities of the Escherichia coli strains. Bioinformatics, 34(13):i89–i95, 06 2018. ISSN 1367-4803. doi: 10.1093/bioinformatics/bty276. URL https://doi.org/10.1093/bioinformatics/bty276.

[71] J.H. Paton and D.S. Reeves. Fluoroquinolone antibiotics. Drugs, 36(2):193–228, August 1988. doi: 10.2165/00003495-198836020-00004. URL https://doi.org/10.2165/00003495-198836020-00004.

[72] F Lindberg and S Normark. Contribution of chromosomal beta-lactamases to beta-lactam resistance in Enterobacteria. Reviews of infectious diseases, 8:S292–S304, 1986. ISSN 0162-0886. doi: 10.1093/clinids/8.Supplement_3.S292.

[73] Li Xiao, Donna M Crabb, Lynn B Duffy, Vanya Paralanov, John I Glass, and Ken B Waites. Chromosomal mutations responsible for fluoroquinolone resistance in Ureaplasma species in the United States. Antimicrobial agents and chemotherapy, 56(5):2780–2783, 2012. doi: 10.1128/AAC.06342-11.

[74] Srikumar S. Deekshit VK. 'to be, or not to be'—the dilemma of 'silent' antimicrobial resistance genes in bacteria. Journal of Applied Bacteriology, 2022. ISSN 1364-5072. doi: 10.1111/jam.15738.

[75] Jee In Kim, Finlay Maguire, Kara K Tsang, Theodore Gouliouris, Sharon J Peacock, Tim A McAllister, Andrew G McArthur, and Robert G Beiko. Machine learning for antimicrobial resistance prediction: Current practice, limitations, and clinical perspective. Clinical microbiology reviews, pages e0017921–e0017921, 2022. ISSN 0893-8512. doi: 10.1128/cmr.00179-21.