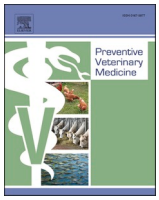
Preventive Veterinary Medicine 246 (2026) 106713

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed



Analyzing multidrug resistance patterns across the food supply chain using association rule mining

Joshua C. Glass [a,*], Gayatri Anil [a,b], Jordan D. Zehr [b], Kristina M. Ceres [b], Laura B. Goodman [b], Casey L. Cazer [a,b]

[a] Department of Clinical Sciences, College of Veterinary Medicine, Cornell University, United States
[b] Department of Public and Ecosystem Health, College of Veterinary Medicine, Cornell University, United States

ARTICLE INFO

Keywords:
Multidrug resistance
Machine learning
One health
Food safety
Surveillance
Escherichia coli

ABSTRACT

We used the machine learning method association rule mining to analyze multidrug resistance (MDR) among cattle-associated *Escherichia coli* along the food supply chain in the USA. All datasets were stratified by year, source, and resistance indicator (genotypic/phenotypic). Pruned rulesets were compared by calculating the proportion of rules from a comparison ruleset that are captured in a reference ruleset. Rulesets were compared across years within each source and indicator type to quantify how MDR patterns change over time. At the class level, on average nearly 50 % or more of the MDR patterns remain the same year over year for genotypic and phenotypic indicators. Rulesets were compared between data sources to quantify how MDR patterns change across the food supply chain. These comparisons suggest that there is a greater diversity of MDR patterns present at slaughterhouse settings than at retail settings; and further, that there is a greater diversity of MDR patterns amongst sick cattle on farm settings than at either slaughterhouse or retail settings. Genetic evidence supports this being attributable to a greater genetic diversity associated with pathogenic bacteria vs commensals. Rulesets were compared between indicators to quantify the degree of correspondence between phenotypic and genotypic data. Genotypic rulesets were better able to capture phenotypic rulesets than the reverse. Adding another aminoglycoside (streptomycin) to the phenotypic analysis, improved ruleset correspondence. This asymmetry may be driven by drug specific aminoglycoside resistance genes, suggesting that more drugs need to be assessed to have a fuller understanding of the variation in MDR patterns.

1. Introduction

Agricultural antimicrobial use selects for drug resistant microbes among food animal populations, which may persist at subsequent levels of the food supply chain, leading to notable impacts on food safety and human health (Marshall and Levy, 2011). Of special importance is the emergence of multi-drug resistance (MDR), typically defined as resistance to three or more antimicrobial classes (Magiorakos et al., 2012). MDR poses a unique challenge to health care professionals by limiting or eliminating treatment options for animal and human illness. To effectively combat the threat of MDR, a cooperative effort must be made involving detailed surveillance and research to provide data driven insights that can help shape antimicrobial stewardship policies and practices.

However, it is challenging to derive clear insights on MDR from antimicrobial resistance surveillance data due to MDR's combinatorial complexity. Given the number of antimicrobials that are commonly included in surveillance efforts, the number of possible unique MDR patterns quickly becomes difficult to manage. Several analysis methods have been applied to MDR data, but many are limited in their ability to account for higher order interactions and/or by unmet statistical assumptions (Love et al., 2018, 2016; Ludwig et al., 2013; Wagner et al., 2003; Zawack et al., 2018, 2016). Unfortunately, the higher order interactions among MDR data are likely the most important aspect of the data. Characterizing relations among resistance patterns is likely to uncover subtle but important findings that could not be understood through simply examining resistance trends for individual antimicrobials.

The associations among MDR patterns and how they change over time and across the food chain is likely reflective of the driving forces behind the emergence of such MDR patterns: there is good reason to believe that the emergence of each of the resistances in an MDR pattern

* Correspondence to: Department of Clinical Sciences, Cornell University College of Veterinary Medicine, Ithaca, NY 14850, United States.
E-mail address: jg2527@cornell.edu (J.C. Glass).

https://doi.org/10.1016/j.prevetmed.2025.106713
Received 26 March 2025; Received in revised form 18 September 2025; Accepted 6 October 2025
Available online 6 October 2025

are not independent of each other (Chang et al., 2015). For example, a set of highly associated resistance patterns may be highly associated because of co-resistance (i.e., a common resistance mechanism), and understanding such associations may reveal an unrecognized common mechanism or highlight a known but under-prioritized mechanism. Additionally, examining the associations among MDR patterns could uncover sources of cross-resistance (i.e., genomically linked resistance mechanisms): a particular bacterial isolate may be resistant to a certain set of structurally unrelated antimicrobials because environmental factors and selective pressures may facilitate the proliferation and persistence of each of the resistances, making it important to understand such environmental factors (for example see Khachatryan et al., 2006); or the genes/mutations conferring these resistances may exist together on the same operon, transposon, or plasmid (Marshall and Levy, 2011; Stokes and Gillings, 2011). Lacking such knowledge could negatively impact the effectiveness of antimicrobial stewardship interventions. For example, a certain antimicrobial could be prohibited for agricultural use to decrease the selective pressure toward resistance for that antimicrobial. If the gene that confers resistance to this antimicrobial is located on the same genetic element as the gene that confers resistance to a different (seemingly less important) antimicrobial that is still in use, then the selective pressures for this MDR pattern would likely persist despite the stewardship efforts.

To overcome the challenges of combinatorial complexity, previous studies have applied association rule mining to the analysis of MDR (Cazer et al., 2021, 2019). Association rule mining is a machine learning technique traditionally used to analyze consumer purchasing behavior. In association rule mining, the data consists of transactions that contain items. The set of transactions are analyzed to find frequent item sets, and from these association rules are constructed. Association rules describe which items are likely to co-occur and are typically conceptualized as the presence of one set of items implying the presence of another set of items. For example, if **A**, **B**, and **C** are items, then a rule could be expressed as {**A, B**} →{**C**}, and would be interpreted as the presence of item set {**A, B**} implies the presence of item set {**C**} (for which there are several numerical measures to quantify the strength of this association). The item sets involved in the rule are often referred to as the left-hand side (LHS) and the right-hand side (RHS) of the rule. Association rule mining manages the computational complexity of the analysis by employing a computationally efficient algorithm (commonly, the Apriori algorithm) that generates and prunes frequent item sets and rules (Agrawal et al., 1993).

While more modern machine learning approaches such as support vector machines (SVM) and deep neural networks (DNN) have demonstrated strong predictive performance for antimicrobial resistance, their 'black box' nature limits their utility for understanding the underlying patterns and mechanisms of resistance (Mosaddegh et al., 2024). Moreover, while these complex predictive models (such as DNNs) require training data to learn from, our goal is to understand patterns in retrospective data through unsupervised, descriptive analysis. Association rule mining is particularly well-suited for this purpose. In contrast to supervised learning methods that are trained to make predictions, association rule mining is an unsupervised approach that discovers and describes inherent patterns in the data, making it an ideal tool for exploring and understanding complex relationships in MDR patterns. Additionally, association rule mining provides transparent, interpretable results that can reveal meaningful biological relationships. Although other interpretable methods exist, such as hierarchical clustering, they often lack the deterministic quality needed for reliable comparisons across different datasets or time points (Mosaddegh et al., 2024). Association rule mining overcomes this limitation while maintaining interpretability, making it particularly well-suited for analyzing how resistance patterns change across time and context. The method has been successfully employed across various antimicrobial resistance contexts. For example, Giannopoulou et al. (2007) developed a nationwide surveillance framework using association rules to detect emerging resistance patterns across Greek hospitals. Zaharia et al. (2019) demonstrated the method's utility in discovering AMR patterns in human *E. coli* isolates, while Liang et al. (2021) applied a weighted version of the algorithm to analyze MDR in bovine mastitis pathogens. The method has been applied to infection control surveillance (Brossette et al., 2000; Ma et al., 2003) and has been shown to complement other analytical approaches such as Bayesian network models (Zaharia et al., 2021).

Two of our previous studies have applied association rule mining to the analysis of MDR by considering each bacterial isolate as a transaction, and each antimicrobial that the isolate was resistant to was considered an item present in that transaction. One study analyzed data from the National Antimicrobial Resistance Monitoring System (NARMS) (Food and Drug Administration (FDA), 2024) for *Escherichia coli* isolates associated with chicken (Cazer et al., 2019), and another study analyzed *Staphylococcus aureus* isolates collected from a hospital setting (Cazer et al., 2021). Both studies applied the association rule mining method to resistance data based on phenotypic indicators of resistance: minimum inhibitory concentrations (MICs) generated from microbiological assays with reference to interpretative breakpoints were used to classify each isolate as resistant or susceptible to each antimicrobial in question. However, these and other previous studies have focused solely on phenotypic resistance data. The current study presents a novel extension by applying association rule mining to both phenotypic and genotypic indicators of resistance. By analyzing both types of data, we can provide a more comprehensive understanding of resistance patterns and their underlying mechanisms. Resistance data from *E. coli* associated with cattle were analyzed using association rule mining and subsequent ruleset comparison metrics to determine MDR trends over time, MDR trends across the food chain, and to investigate the relationship between phenotypic and genotypic indicators of resistance.

2. Methods

2.1. Data preparation

Three datasets were analyzed in total: data from retail meat samples retrieved from NARMS (Food and Drug Administration (FDA), 2024), data from cecal (slaughterhouse) samples retrieved from NARMS, and data from clinically ill animals retrieved from the National Animal Health Laboratory Network (NAHLN) (National Animal Health Laboratory Network, 2023). The NARMS program is a national antimicrobial resistance surveillance program collecting samples from slaughterhouses and retail settings, focused on foodborne pathogens ("The National Antimicrobial Resistance Monitoring System Strategic Plan, 2021-2025, 2021–2025 2021–2025," 2021). The NAHLN antimicrobial resistance program collects data from participating veterinary diagnostic laboratories on bacteria isolated from animal samples submitted as clinical cases (USDA APHIS VS National Animal Health Laboratory Network (NAHLN) Antimicrobial Resistance Pilot Project, 2020). For all datasets, only *E. coli* isolates associated with cattle were examined. The retail meat dataset included minimum inhibitory concentrations (MICs) for years 2002–2021 (2020 phenotypic data was not available), and genotypic information for years 2017–2021. The cecal dataset included MICs for years 2013–2023, and genotypic information for years 2017–2021 (genotypic information for 2019 was not available for the cecal dataset). The sick cattle dataset included MICs and genotypic information for years 2018–2022.

All genomes used in this study were downloaded via NCBI's datasets tool (v. 16.9.0, https://www.ncbi.nlm.nih.gov/datasets/docs/v2/command-line-tools/download-and-install/). We used GNU parallel v.20170522 (Tange, 2015) to run AMRFinderPlus v.4.0.3 (Feldgarden et al., 2021; https://github.com/ncbi/amr) with database v.2024.10.22.1 with the full search combination (-p, -g, -n, –organism Escherichia, –plus) to screen each assembly for the presence of antimicrobial genes and point mutations. PlasmidHunter v.1.4 (Tian et al.,

2024; https://github.com/tianrenmaogithub/PlasmidHunter) was used to screen each assembly for plasmids. AMRFinder has been shown to be highly accurate on NARMS isolates when compared to other gene-calling tools and phenotypic traits (Feldgarden et al., 2019). All genomes used in this study can be found in Suppl. Table 1.

MIC data were examined for the drugs listed in Table 1 and were interpretated as resistant or susceptible based on breakpoints from NARMS interpretive criteria or epidemiologic cut-off values (ECOFFS) retrieved from EUCAST (European Committee on Antimicrobial Susceptibility Testing, 2024; Food and Drug Administration, 2021). The genotypic information for each dataset was comprised of resistance genes as well as genetic mutations that were included due to their likelihood of conferring resistance (mistranslations were not considered).

The goals of this study were to examine how multidrug resistance (MDR) patterns change over time, how MDR patterns change across the food supply chain, and how well MDR patterns as indicated by phenotypic data correspond to those indicated by genotypic data. All datasets were stratified to produce the appropriate comparison sets. The three original datasets represent three different points in the food supply chain: comparing across these three data sources achieves the goal of quantifying change across the food chain. The three datasets were further divided based on year to quantify change across time. Finally, all datasets were divided based on resistance indicator (i.e., genotypic or phenotypic). Each source-year-indicator dataset was analyzed separately. Prevalence descriptives for phenotypic drug resistance and resistance genes/mutations can be found in Tables 2–4.

2.2. Association mining

Association rule mining was implemented use the R package arules (Hahsler et al., 2023, 2011, 2005). To appropriately format the data for association rule mining, each bacterial isolate was treated as a transaction. When mining phenotypic data, the set of drugs tested were treated as the possible items for the transaction: if the isolate was resistant to a drug, it was coded as 1 (the item is present in the transaction); and if the isolate was susceptible to a drug, it was coded as 0 (the item is not present in the transaction). Similarly, when mining genotypic data, the set of possible genes/mutations were treated as the items for the transaction: if a gene/mutation was present in the isolate, it was coded as 1; and if a gene/mutation was not present in the isolate it was coded as 0. The data were mined both at the individual item level (mining drugs or genes/mutations) and at the antimicrobial class level. For phenotypic data each drug was mapped to its corresponding class. If multiple drugs mapped to the same class, the largest value for that class for each isolate was retained (for example, if three drugs of the same class were coded as [0,0,1] for a particular isolate, then the final coding for the class would be 1 for that isolate). The same procedure was followed for the resistance genes/mutations. Classifications for genes/mutations were retrieved from the NCBI Bacterial Antimicrobial Resistance Reference Gene Database v. 2024–12–18.1 ("Reference Gene Catalog," n.d., https://www.ncbi.nlm.nih.gov/pathogens/refgene/). In the case where a single gene conferred resistance to multiple classes of antimicrobials, an item for each of the classes was created with identical codings from that gene (for example if a gene that confers resistance to two classes was coded as 1, then both classes would be included and would be coded as 1).

2.2.1. Quality measure selection

If all possible rules are generated, it is likely that some will be spurious, and there may be some that are not spurious but that are uninformative. Therefore, it is important to be able to prune rules based on their quality (i.e., how interesting they are and how likely they are to describe meaningful associations), and there are numerous measures of rule quality available (Michael Hahsler, 2015).

To determine which quality measure would be most appropriate for

Table 1
Drug Information. Breakpoints with * indicate that these are ECOFFs from EUCAST; all other breakpoints are from NARMS interpretative criteria. NB = No Breakpoints.

Class	Drug	Abbreviation	Resistant Breakpoint	In Dataset
AMINOGLYCOSIDE	Gentamicin	GEN	> =16	Retail meat; Cecal; Sick cattle
	Streptomycin	STR	> =32	Retail meat; Cecal
	Neomycin	NEO	> 8*	Sick cattle
	Spectinomycin	SPE	> 64*	Sick cattle
BETA-LACTAM	Amoxicillin–Clavulanic Acid	AMC	> =32	Retail meat; Cecal
	Meropenem	MER	> =4	Retail meat; Cecal
	Cefoxitin	FOX	> =32	Retail meat; Cecal
	Ceftriaxone	AXO	> =4	Retail meat; Cecal
	Ampicillin	AMP	> =32	Retail meat; Cecal; Sick cattle
	Ceftiofur	TIO	> =8	Sick cattle
	Penicillin	PEN	NB	Sick cattle
FOLATE-PATHWAY-INHIBITOR	Sulfisoxazole	FIS	> 256	Retail meat; Cecal
	Sulfamethoxazole	SMX	> 256	Retail meat
	Trimethoprim/sulfamethoxazole	COT	> =4	Retail meat; Cecal; Sick cattle
	Sulphadimethoxine	SUL	NB	Sick cattle
LINCOSAMIDE	Clindamycin	CLI	NB	Sick cattle
MACROLIDE	Azithromycin	AZI	> =32	Retail meat; Cecal
	Gamithromycin	GAM	NB	Sick cattle
	Tildipirosin	DIP	NB	Sick cattle
	Tilmicosin	TIL	NB	Sick cattle
	Tulathromycin	TUL	NB	Sick cattle
	Tylosin	TYL	NB	Sick cattle
PHENICOL	Chloramphenicol	CHL	> =32	Retail meat; Cecal
	Florfenicol	FFN	> 16*	Sick cattle
PLEUROMUTILIN	Tiamulin	TIA	NB	Sick cattle
QUINOLONE	Ciprofloxacin	CIP	> =1	Retail meat; Cecal
	Nalidixic acid	NAL	> =32	Retail meat; Cecal
	Danofloxacin	DAN	NB	Sick cattle
	Enrofloxacin	ENR	> 0.125*	Sick cattle
TETRACYCLINE	Tetracycline	TET	> =16	Retail meat; Cecal; Sick cattle
	Chlortetracycline	LOR	NB	Sick cattle
	Oxytetracycline	OXY	NB	Sick cattle

Table 2
Phenotypic Resistance Prevalence (%) Descriptives for Retail Meat and Cecal Datasets. NT = Not Tested. N is the number of isolates. Antimicrobial abbreviations are in Table 1.

	Year	N	AMC	AMP	AXO	AZI	CHL	CIP	COT	FIS	FOX	GEN	MER	NAL	SMX	TET
Retail Meat	2002	295	2.03	6.1	0	NT	1.02	0	0.68	NT	0	0.34	NT	0	10.17	30.85
	2003	311	2.25	5.14	0.32	NT	2.25	0	0.32	NT	0	0.96	NT	0.96	10.29	25.08
	2004	338	3.85	5.33	1.48	NT	3.55	0	0.59	13.02	1.18	0.59	NT	1.48	NT	22.78
	2005	316	1.27	3.48	1.9	NT	1.58	0.32	0.63	6.96	0.95	0	NT	1.27	NT	16.46
	2006	295	2.37	9.15	1.69	NT	1.36	0	1.36	12.54	2.03	4.07	NT	0.68	NT	25.42
	2007	256	0.78	6.64	0.78	NT	3.91	0	1.17	9.38	0.78	0	NT	0.39	NT	21.88
	2008	250	2.4	6.4	1.6	NT	0.8	0	2	11.6	2.4	2	NT	0.4	NT	24
	2009	247	1.62	4.88	0.81	NT	2.43	0	2.02	7.69	1.62	0.81	NT	0.41	NT	18.62
	2010	269	1.12	4.83	1.12	NT	2.6	0	0.74	12.64	1.12	0.37	NT	0	NT	22.68
	2011	215	0.47	3.72	0.47	0	1.4	0	2.33	7.91	0.47	0.47	NT	0	NT	17.67
	2012	271	1.48	2.58	0	0	1.11	0	0.37	7.38	1.85	0.74	NT	1.48	NT	22.14
	2013	227	1.76	4.85	2.2	0	3.96	0	1.76	7.93	1.32	0	NT	0.44	NT	22.47
	2014	205	0.49	4.39	0.49	0	0.49	0	0.98	7.8	0.98	0.49	NT	0	NT	21.46
	2015	227	0.44	2.64	0.44	0	1.32	0	1.32	7.49	0.44	0.44	NT	0.44	NT	18.5
	2016	174	1.15	6.32	0.57	0	5.17	0	1.15	9.77	1.15	0.57	0	0	NT	23.56
	2017	271	0.37	6.27	0.37	0	2.58	0	1.11	6.27	0	0	0	1.85	NT	21.77
	2018	310	2.26	9.35	1.94	0.65	7.1	0.32	6.77	14.52	1.29	1.61	0	2.27	NT	27.1
	2019	286	0	2.1	0	0	2.45	0	0.7	6.64	0	0.7	0	0.7	NT	19.58
	2020	0	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT
	2021	386	1.3	6.74	1.04	0.26	4.15	0.26	2.07	10.62	0.78	1.04	0	1.04	NT	20.98
	Mean	5149	1.44	5.31	0.91	0.09	2.59	0.05	1.48	9.42	0.97	0.8	0	0.73	10.23	22.26
	Standard Deviation	5149	0.95	1.98	0.71	0.21	1.7	0.11	1.42	2.59	0.71	0.95	0	0.68	0.08	3.43
Cecal	2013	293	0	3.75	0	0	3.75	0	0	9.22	0	0.34	NT	0	NT	25.94
	2014	326	0.61	5.21	0.92	0	3.68	0	0.92	11.96	0.61	0.61	NT	0.31	NT	37.42
	2015	566	1.24	6.54	0.88	0	6.36	1.24	0.53	16.43	1.24	0.71	NT	2.13	NT	39.22
	2016	738	0.81	6.1	0.81	0.27	6.78	0.41	0.54	14.63	0.68	0	0	1.09	NT	38.75
	2017	936	0.43	5.02	0.53	0.11	5.24	0	0.75	15.71	0.32	0	0	0.97	NT	37.61
	2018	1029	0.49	6.51	0.39	0	8.07	0.39	0.87	15.45	0.49	0.29	0	1.86	NT	35.67
	2019	759	0.4	5.67	0.4	0.13	6.72	0.13	1.32	13.57	0.26	0.26	0	1.86	NT	33.2
	2020	477	6.5	23.9	7.34	1.05	14.47	1.26	12.58	25.58	6.71	1.89	0	2.1	NT	44.03
	2021	282	2.48	13.12	5.32	1.42	10.64	0.71	6.03	19.86	2.84	2.13	0	1.79	NT	39.36
	2022	407	6.14	22.85	8.6	1.47	16.22	0.49	14	28.75	5.9	1.47	0	1.72	NT	44.96
	2023	219	0.46	8.22	0.46	0	5.94	0.91	1.83	15.07	0.46	0.46	0	1.37	NT	32.88
	Mean	6032	1.78	9.72	2.33	0.4	7.99	0.5	3.58	16.93	1.77	0.74	0	1.38	NT	37.19
	Standard Deviation	6032	2.34	7.18	3.15	0.6	4.13	0.47	5.07	5.76	2.37	0.75	0	0.72	NT	5.31

the data, rules were mined for each source-year-indicator dataset, and all available quality measures were calculated. Principal component analysis was performed on the values of all available quality measures, and the first four principal components were examined. PCA was implemented using the prcomp function, which is included in base R (R Core Team, 2024). In each of the first four principal components the top five variable loadings were extracted for each source-year-indicator dataset, and the frequency with which each measure was present in the top five variable loadings was recorded for each principal component.

In addition to this, the mathematical properties of the measures were considered. To appropriately evaluate association rules mined from sparse data (such as resistance data), the quality measures used should have two properties: null-invariance (the strength of the association is influenced by the presence of items rather than their absence) and symmetry (the value of the measure is the same for A→B and B→A) (Tan et al., 2002). From the measures that were found in the PCA procedure, four null invariant and symmetrical measures were chosen: cosine, Jaccard, Kulczynski, and support. Support is a commonly used quality measure and simply reports how often a rule appears as a proportion of the sample size of the dataset. The formula for support can be seen in Eq. 1.

$$\boldsymbol{supp}(\boldsymbol{X}) = \frac{\boldsymbol{nX}}{\boldsymbol{n}} = \boldsymbol{P}(\boldsymbol{X}) \tag{1}$$

Kulczynski is a modified version of confidence. Confidence represents the conditional probability that given the left-hand side of the rule, the right-hand side of the rule will be present as well. Confidence is not symmetrical. Kulczynski is made symmetrical by taking both directions of the conditional probability and averaging them. Kulczynski ranges from 0 to 1, with 0 meaning that the RHS and LHS never occur together, 1 meaning that they always occur together, and 0.5 suggesting that the rule is probably uninteresting (Eq. 2).

$$\mathbf{kulc}(\mathbf{X}\rightarrow\boldsymbol{Y}) = \frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{P}(\boldsymbol{X}|\boldsymbol{Y}) + \boldsymbol{P}(\boldsymbol{Y}|\boldsymbol{X})) \tag{2}$$

Jaccard shows how strongly the RHS and LHS are associated by dividing the proportion of transactions where the RHS and LHS co-occur by the proportion of total occurrences of the RHS or LHS. Jaccard ranges from 0 to 1, and a Jaccard value of 0.33 suggests that the RHS and LHS occur together at chance levels (Eq. 3).

$$\boldsymbol{jaccard}(\mathbf{X}\rightarrow\boldsymbol{Y}) = \frac{\boldsymbol{supp}(\boldsymbol{X}\cap\boldsymbol{Y})}{\boldsymbol{supp}(\boldsymbol{X}) + \boldsymbol{supp}(\boldsymbol{Y}) - \boldsymbol{supp}(\boldsymbol{X}\cup\boldsymbol{Y})} \tag{3}$$

Cosine is a correlation measure similar to the lift measure. Cosine was chosen instead of lift because of its null invariance property. Cosine ranges from 0 to 1; 0.5 suggests that there is no correlation between the RHS and LHS, values close to zero suggest a negative correlation, and values close to 1 suggest a positive correlation.

$$\boldsymbol{cosine}(\mathbf{X}\rightarrow\boldsymbol{Y}) = \frac{\boldsymbol{supp}(\boldsymbol{X}\cap\boldsymbol{Y})}{\sqrt{(\boldsymbol{supp}(\boldsymbol{X})\boldsymbol{supp}(\boldsymbol{Y}))}} \tag{4}$$

2.2.2. Cut-off selection

Once appropriate quality measures were chosen, the value of these measures to apply as a cut-off threshold for pruning rules needed to be set. The selected measures have values that could be considered intrinsically meaningful cut-off values (i.e., a cosine value of 0.5 to retain only rules that represent a positive correlation); additionally, the number of rules needed to be retained to fully describe the dataset was considered in the setting of cut-off values. To determine how many rules are needed to fully describe the data, random subsamples of rules were

Table 3
Phenotypic Resistance Prevalence (%) Descriptives for the sick cattle dataset. NB = No Breakpoints. NI = Not Interpretable due to insufficient concentration range tested. N is the number of isolates. Antimicrobial abbreviations are in Table 1.

	Year	N	AMP	CLI	COT	DAN	DIP	ENR	FFN	GAM	GEN	LOR	NEO	OXY	PEN	SPE	SUL	TET	TIA	TIL	TIO	TUL	TYL
Sick cattle	2018	406	NI	NB	13.79	NB	NB	38.92	NI	NB	15.76	NB	34.73	NB	NB	33.00	NB	NI	NB	NB	33.50	NB	NB
	2019	635	NI	NB	2.68	NB	NB	32.28	NI	NB	11.97	NB	35.12	NB	NB	30.24	NB	NI	NB	NB	26.77	NB	NB
	2020	481	NI	NB	1.66	NB	NB	34.3	NI	NB	11.64	NB	35.14	NB	NB	33.47	NB	NI	NB	NB	32.22	NB	NB
	2021	530	NI	NB	2.08	NB	NB	28.68	NI	NB	11.32	NB	27.74	NB	NB	31.51	NB	NI	NB	NB	28.87	NB	NB
	2022	419	NI	NB	2.39	NB	NB	23.87	NI	NB	9.55	NB	26.25	NB	NB	24.34	NB	NI	NB	NB	26.97	NB	NB
	Mean	2471	NI	NB	4.52	NB	NB	31.61	NI	NB	12.05	NB	31.8	NB	NB	30.51	NB	NI	NB	NB	29.67	NB	NB
	Standard Deviation	2471	NI	NB	5.2	NB	NB	5.69	NI	NB	2.28	NB	4.42	NB	NB	3.68	NB	NI	NB	NB	3.06	NB	NB

taken at various subsample sizes and the number of antimicrobial classes represented by each subsample was computed (Fig. 1A). This determined the minimum number of rules required to ensure that all the antimicrobial classes were represented in the ruleset. To select cut-off values for each quality measure, rules were mined and pruned at various cut-off values. Cut-off values that achieved the target rule number were chosen (Fig. 1B-E). The cut-off value for support was set at 0 to ensure that rare but important resistance patterns were able to be found. Cosine and kulczynski were both set at 0.5, which was found to prune the desired quantity of rules in the majority of datasets, and which was in line with both measures' intrinsically meaningful cut-off values. The cut-off value for Jaccard was set to 0 because it was found that increasing this value resulted in over-pruning in a notable number of the datasets when pruning jointly with the other selected measures; (note that Fig. 1A-D show pruning for each measure independently, rather than the combined effect of applying the cut-offs jointly). Phenotypic datasets derived from the retail meat samples were particularly susceptible to over-pruning likely due to lower sample sizes and, possibly, more sparse data; consequently, these datasets generated a smaller average quantity of pre-pruned rules. To maintain consistency for later ruleset comparison, the same cut-off values were applied to all datasets, which unavoidably constrained the strictness of the cut-off values to those values found to be most appropriate for the datasets with the least number of pre-pruned rules (i.e., the phenotypic retail meat datasets).

2.3. Ruleset analysis methods

2.3.1. Proportion captured metric

This analysis was intended to achieve three general goals: to determine MDR trends across time, to determine MDR trends across various sources in the food supply chain, and to quantify the degree of correspondence between phenotypic and genotypic indicators of MDR. Previous studies have used metrics of rule overlap or cumulative rule overlap when comparing rulesets (Cazer et al., 2019; Dudek, 2010), in which rule overlap was calculated by dividing the magnitude of the intersection of the two rulesets being compared by the magnitude of the union of the two rulesets. In the cumulative case, the values for each comparison are progressively averaged across time and incorporated into the subsequent calculations.

It was found that these previously used metrics were too coarse to assess ruleset similarity for the current datasets and analysis goals. Due to the characteristics of the data being analyzed, there were differences across datasets that were irrelevant to the analysis goals but that contributed to overestimating ruleset dissimilarity when employing the previously used rule overlap metrics. For example, when comparing two rulesets from two consecutive years, a particular drug may have been evaluated in one year but not in the other. Any rule that contains this drug will never be included in the rule overlap; however, it is not necessarily the case that this difference between rulesets is representative of a meaningful difference in MDR patterns since it is unknown what the rulesets would include had the drug been tested in both years. Another example: when comparing genotypic and phenotypic rulesets, the genotypic ruleset may include all the rules that are contained in the phenotypic ruleset but may contain many more rules that are not present in the phenotypic ruleset (such that the phenotypic ruleset is a subset of the genotypic ruleset). The resulting rule overlap value would likely be relatively low (despite the phenotypic rules being a subset of the genotypic rules) due the large number of additional rules unique to the genotype ruleset. In this case, the most useful information to be extracted from such a comparison is the subset relation between the phenotypic and genotypic rulesets; however, the rule overlap metric would imply only the somewhat misleading conclusion that the two rulesets are dissimilar.

The proportion captured metric was introduced as a more sensitive alternative to rule overlap. The proportion captured metric assesses how well one ruleset can represent another ruleset by dividing the magnitude

Table 4
Genotypic Resistance Prevalence (%) Descriptives at the Antimicrobial Class Level. N is the number of isolates.

	Year	N	AMINOGLYCOSIDE	BETA – LACTAM	BLEOMYCIN	FOLATE – PATHWAY – INHIBITOR	FOSFOMYCIN	FOSMIDOMYCIN	LINCOSAMIDE	MACROLIDE	MULTIDRUG	NITROFURAN	PHENICOL	QUINOLONE	RIFAMYCIN	STREPTOGRAMIN	STREPTOTHRICIN	TETRACYCLINE	TRICLOSAN
Retail Meat	2017	142	9.15	5.63	0.7	5.63	93.66	4.23	0	0.7	1.41	0	2.82	3.52	0	0	2.11	26.06	0
	2018	123	18.7	5.69	0.81	12.2	98.37	4.88	0	0	0.81	0	4.07	3.25	0	0	5.69	24.39	0
	2019	271	12.18	2.95	0	8.12	97.79	1.85	0.37	0	0.74	0.37	2.58	1.48	0	0	3.32	23.62	0
	2020	167	12.57	6.59	0	8.98	95.21	2.99	0	0	0	0	2.4	0	0	0	1.8	20.96	0
	2021	343	15.16	5.54	0	12.24	97.67	3.79	0	0.29	2.33	0	4.66	2.33	0	0	5.83	20.41	0
	Mean	1046	13.55	5.28	0.3	9.43	96.54	3.55	0.07	0.2	1.06	0.07	3.31	2.12	0	0	3.75	23.09	0
	Standard Deviation	1046	3.58	1.37	0.42	2.83	2.02	1.17	0.17	0.31	0.87	0.17	1	1.43	0	0	1.92	2.37	0
Cecal	2017	87	86.21	39.08	2.3	85.06	94.25	3.45	0	1.15	0	1.15	31.03	10.34	0	0	2.3	93.1	0
	2018	409	38.88	17.36	0.98	34.72	94.62	1.96	0	0.24	2.2	0.24	19.07	4.89	0	0	3.67	53.79	0
	2019	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0
	2020	188	60.64	45.74	0.53	53.19	88.83	11.7	3.19	2.66	0	2.66	26.6	7.98	0.53	0	7.45	70.74	0
	2021	165	35.15	21.82	0	33.94	93.33	4.85	1.82	2.42	2.42	0.61	17.58	5.45	0	0.61	4.24	54.55	0
	Mean	849	55.22	31	0.95	51.73	92.76	5.49	1.25	1.62	1.15	1.16	23.57	7.16	0.13	0.15	4.42	68.04	0
	Standard Deviation	849	23.52	13.58	0.98	23.94	2.67	4.3	1.55	1.13	1.34	1.06	6.35	2.51	0.26	0.3	2.18	18.44	0
Sick cattle	2018	31	54.84	41.94	0	54.84	93.55	6.45	0	0	0	9.68	38.71	32.26	0	0	0	67.74	0
	2019	74	59.46	56.76	2.7	56.76	85.14	12.16	9.46	6.76	0	0	37.84	27.03	1.35	0	5.41	71.62	1.35
	2020	83	54.22	46.99	1.2	51.81	85.54	9.64	2.41	1.2	3.61	1.2	31.33	13.25	0	0	4.82	59.04	0
	2021	91	56.04	39.56	1.1	52.75	90.11	5.49	8.79	10.99	1.1	1.1	40.66	26.37	3.3	2.2	2.2	65.93	0
	2022	3	33.33	33.33	0	33.33	100	0	0	0	0	0	33.33	0	0	0	0	33.33	0
	Mean	282	51.58	43.72	1	49.9	90.87	6.75	4.13	3.79	0.94	2.4	36.37	19.78	0.93	0.44	2.49	59.53	0.27
	Standard Deviation	282	10.4	8.79	1.11	9.46	6.17	4.61	4.67	4.9	1.57	4.11	3.89	13.09	1.45	0.98	2.57	15.34	0.6

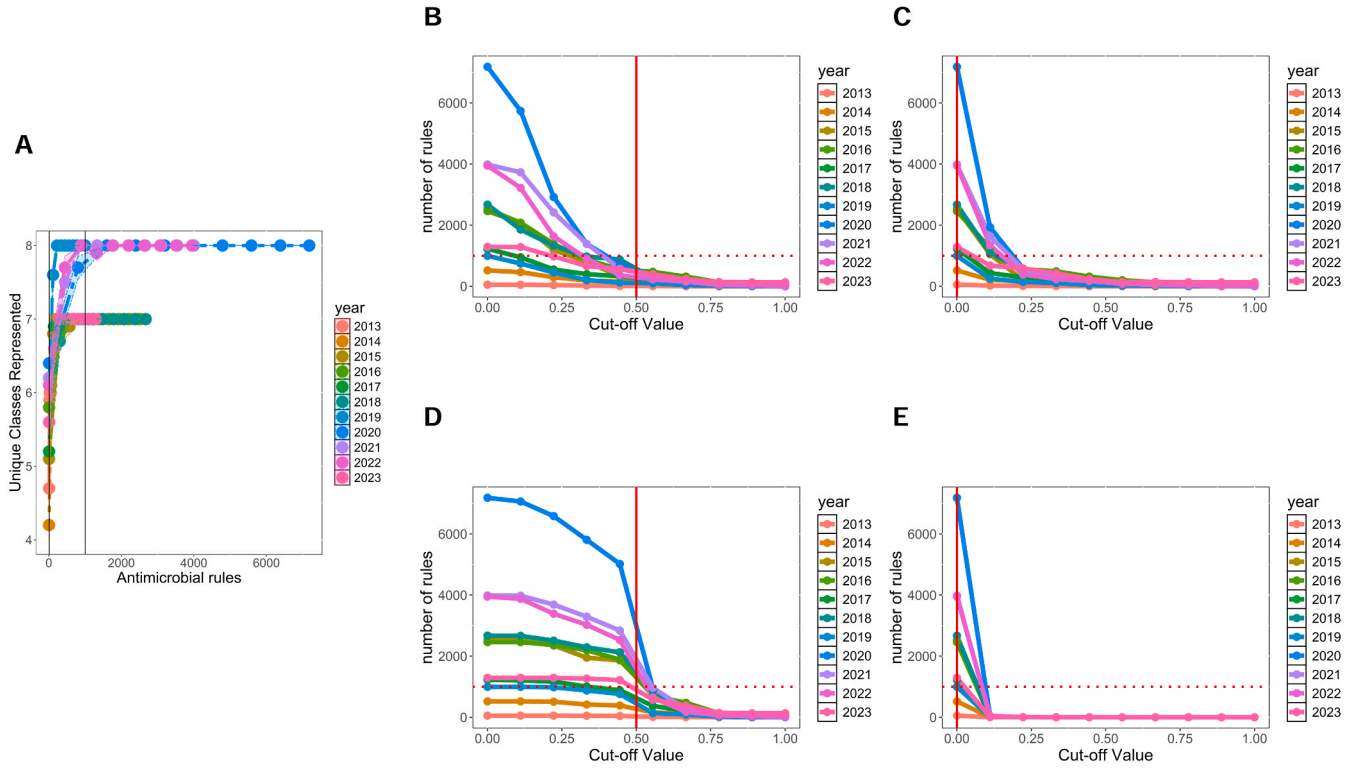


Fig. 1. Cut-off selection example. Plot A shows subsamples of rules plotted against the number of antimicrobial classes represented by those rules. The number of rules remaining after pruning with a sample of different cut-offs for cosine (B), jaccard (C), kulczynski (D), and support (E). The final chosen cut-off values for each measure are indicated by a vertical line in each plot. Target rule number is indicated by a horizontal dotted line.

of the intersection of the two rulesets being compared by the magnitude of only the first ruleset (Eq. 5). For terminological clarity, the ruleset "trying" to capture the information in another ruleset (R1 in Eq. 5) will be called the reference ruleset, and the ruleset "being captured" (R2 in Eq. 5) will be called the comparison ruleset. Quantifying how well a reference ruleset can capture a comparison ruleset, lessens the impact of irrelevant dataset differences as compared to quantifying exact (i.e., bidirectional) correspondence. It should be noted that the rule overlap metric was symmetrical, but the proportion captured metric is asymmetrical: if the reference ruleset and comparison ruleset are reversed, the results may differ. This asymmetry adds more fine-grained information on the relation between the two rulesets. In addition to lessening the impact of irrelevant dataset differences, the proportion captured metric can provide more explicit insights into the relative information content of the rulesets being analyzed. For example, when comparing phenotypic and genotypic rulesets, the proportion captured metric can be calculated in both directions (i.e., phenotype as reference ruleset with genotype as comparison ruleset, and genotype as reference ruleset with phenotype as comparison ruleset). Joint consideration of these proportion captured values provides the insights that the rule overlap was able to provide along with additional insights unique to the proportion captured metric. If both proportion captured values are high, then the two rulesets are very similar, and if both are low then the two rulesets are very dissimilar. If one proportion captured value is high and the other low, this suggests the presence of a subset relation such as was mentioned in the earlier example.

To accommodate the temporal nature of the data and to provide the most interpretable level of summary information, the proportion captured metric was applied cumulatively as cumulative proportion captured (CPC; Eqs. (6) and (7)). To accomplish the first analysis goal (to determine how MDR patterns change across time), comparisons were made between datasets that were of the same source and resistance indicator type but of different years. Years were compared consecutively according to Eq. (6). Noting again the asymmetrical quality of the proportion captured metric, when comparing two consecutive years it would be possible to compare them in either direction. However, for two consecutive years, the previous year was always the reference set, and the subsequent year was always the comparison set. This direction of comparison was chosen because it provides more interpretable insights into how MDR patterns evolve over time, showing how well historic patterns represent those that emerge in subsequent years. To accomplish the second analysis goal (to quantify how MDR patterns change across the food supply chain), comparisons were made between datasets that were of the same resistance indicator type and year but of a different source. These calculations were computed according to Eq. (7) in which R1 is a ruleset mined from one source and R2 is a ruleset mined from a different source. For these between sources comparisons, CPC was calculated in both directions. To accomplish the third analysis goal (to quantify how well genotypic and phenotypic indicators of resistance correspond to each other), comparisons were made between datasets that were of the same year and source but of different indicator type. Similarly to the between sources comparisons, the between indicators comparisons were computed according to Eq. (7) in which R1 is a ruleset mined from one resistance indicator type and R2 is a ruleset mined from a different resistance indicator type. Rules were mined at the class level and only classes present in both datasets were retained to make phenotypic and genotypic rules comparable. CPC was calculated in both directions.

To ensure appropriate application of these comparison metrics, datasets were harmonized according to the type of comparison being performed. For comparisons across time (Eq. 6), each dataset was analyzed independently, examining changes between years within each source and resistance indicator type; therefore, no harmonization was needed. For comparisons between sources (Eq. 7), only antimicrobial drugs/classes or genes/gene classes that were present in both datasets being compared were retained prior to association mining, and

comparisons were made only at time points common to both datasets. For comparisons between phenotypic and genotypic indicators (Eq. 7), all data were analyzed at the antimicrobial class level, retaining only classes present in both datasets and comparing only time points for which both types of data were available. Additionally, only isolates with both phenotypic data and corresponding genotypic data available were included in these (phenotypic vs. genotypic) analyses.

$$\textit{Proportion Captured}(\boldsymbol{R1},\quad \boldsymbol{R2}) = \frac{|\boldsymbol{R1} \cap \boldsymbol{R2}|}{|\boldsymbol{R1}|} \tag{5}$$

$$\boldsymbol{CPC}_{\textit{within indicators}} = \begin{cases} \textit{Proportion Captured}(\boldsymbol{R}_{\prime},\ \boldsymbol{R}_{\prime+1}), \textit{if} \prime = 1 \\ \frac{1}{\prime}(\prime-1)\boldsymbol{CPC}_{\prime-1} + \textit{Proportion Captured}(\boldsymbol{R}_{\prime},\ \boldsymbol{R}_{\prime+1}), \textit{if} \prime > 1 \\ \prime = \textit{year index} \end{cases} \tag{6}$$

$$\boldsymbol{CPC}_{\textit{between indicators/sources}} = \begin{cases} \textit{Proportion Captured}(\boldsymbol{R1}_{\prime},\quad \boldsymbol{R2}_{\prime}), \textit{if} \prime = 1 \\ \frac{1}{\prime}(\prime-1)\boldsymbol{CPC}_{\prime-1} + \textit{Proportion Captured}(\boldsymbol{R1}_{\prime},\quad \boldsymbol{R2}_{\prime}), \textit{if} \prime > 1 \\ \prime = \textit{year index} \end{cases} \tag{7}$$

2.3.2. Tabulation comparison

To provide a reference to how well association rule mining functions as a method for investigating MDR, the association rule mining analysis was compared to a more traditional tabulation method of analysis. All unique MDR patterns were found and frequencies for each pattern were tabulated. A pattern present in an isolate was only counted toward the tabulation if it was the complete MDR pattern for that isolate (i.e., containing information from all drugs tested for that isolate or all genetic elements found in that isolate). The tabulated patterns were then compared to the rule data from the association rule mining analysis.

2.4. Virulence gene and plasmid analysis

The average count of virulence genes per isolate within each data source was examined to probe whether the samples taken from each of the data sources were predominantly pathogens or commensal organisms as well as an indicator of the genetic diversity across data sources. Using a Bayesian Poisson regression model with observation level random effects to account for overdispersion (Harrison, 2014) we found substantial differences in counts across data sources. The model was implemented using a Markov chain Monte Carlo algorithm (50,000 iterations, 10,000 burn-in) from the R package MCMCglmm (Hadfield, 2010). Convergence was confirmed by computing Gewecke diagnostics using the coda package (Plummer et al., 2006).

The top 1000 rules from each data source were compared to corresponding plasmid data. Each plasmid combination was matched to its closest rule based on a mismatch distance metric (1-(intersection/union), 0 is a perfect match, 1 is a perfect mismatch). For each plasmid-rule match, the genes in the rule were mapped to the class level and common class combinations were aggregated.

3. Results

3.1. Rule pruning results

The number of pre-pruned and pruned genotypic and phenotypic rules, at the individual item level and class level, for each dataset are in Table 5. In general, there were many more genotypic rules at the individual item level than phenotypic rules. There were substantially fewer class-level rules than individual item level rules, although there were still often 10 times or more genotypic class rules than phenotypic class rules. The retail meat rulesets had the fewest genotypic rules, and the sick cattle rulesets had the most genotypic rules.

3.2. Ruleset comparison

Figs. 2–4 show the results of the ruleset comparisons. Fig. 2 shows the results of temporal comparisons within data sources and resistance indicators. Due to differences in temporal coverage, the time span of

Table 5
Average and standard deviation of the number of rules in pre-pruned and pruned rulesets mined from each of the three data sources. Descriptives for class level rulesets and drug or gene/mutation rulesets are both shown.

	Retail Meats	Cecal	Sick Cattle
Pre-pruned Phenotypic Rules	μ = 1010 (σ = 1476)	μ = 2443 (σ = 2037)	μ = 152 (σ = 55)
Pruned Phenotypic Rules	μ = 195 (σ = 224)	μ = 260 (σ = 145)	μ = 37 (σ = 12)
Pre-pruned Phenotypic Class Rules	μ = 88 (σ = 75)	μ = 180 (σ = 84)	μ = 27 (σ = 7)
Pruned Phenotypic Class Rules	μ = 11 (σ = 9)	μ = 22 (σ = 11).	μ = 11 (σ = 2)
Pre-pruned Genotypic Rules	μ = 1870,032 (σ = 4033,168)	μ = 1098,892 (σ = 1161,887)	μ = 15,259,295 (σ = 14,615,094)
Pruned Genotypic Rules	μ = 1153,139 (σ = 2495,815)	μ = 162,491 (σ = 164,210)	μ = 2284,664 (σ = 2361,652)
Pre-pruned Genotypic Class Rules	μ = 672 (σ = 441)	μ = 3369 (σ = 1821)	μ = 4650 (σ = 4011)
Pruned Genotypic Class Rules	μ = 100 (σ = 59)	μ = 189 (σ = 66)	μ = 590 (σ = 267)

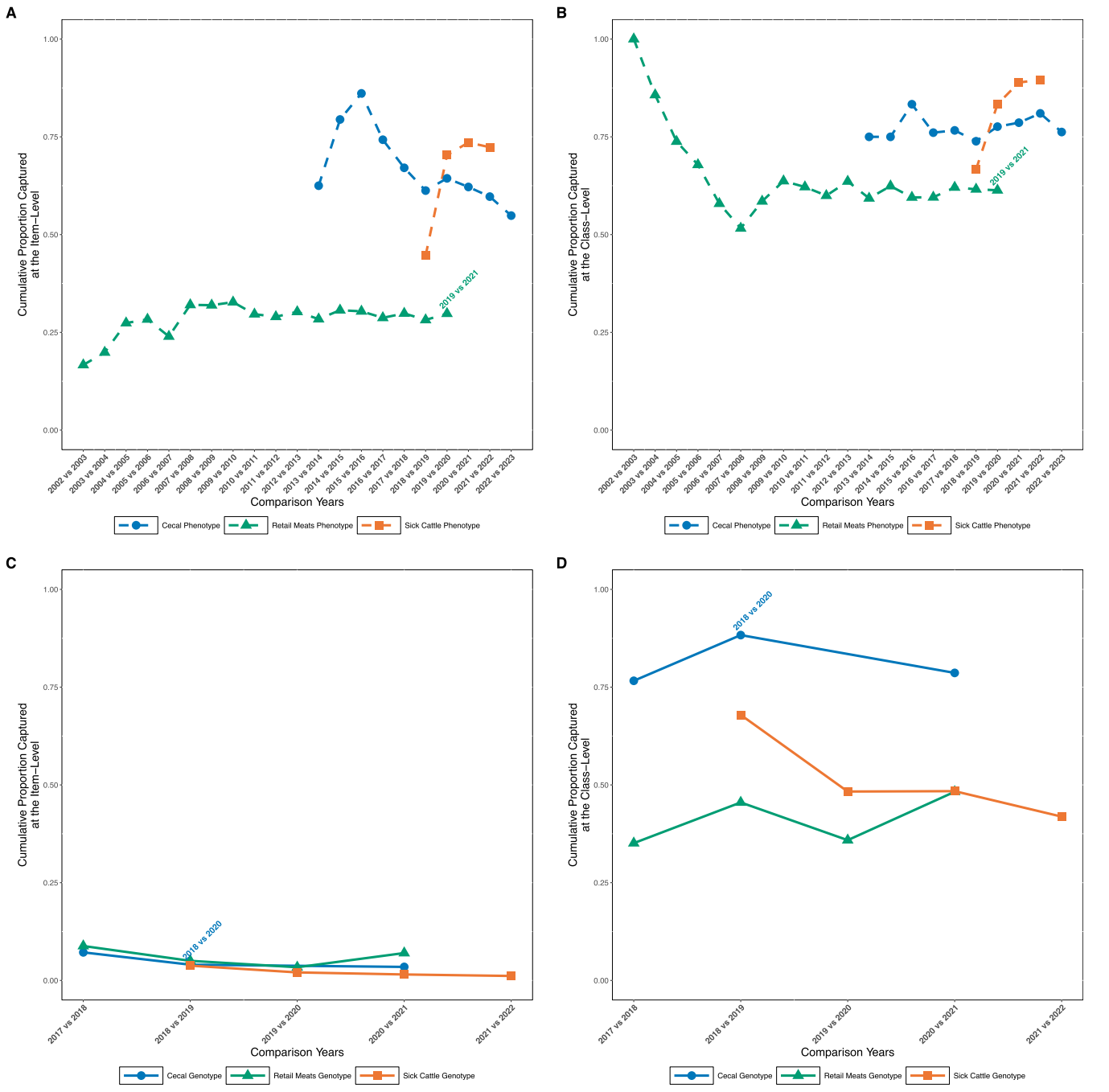


Fig. 2. Temporal comparisons of rulesets within data sources and resistance indicators. Cumulative proportion captured (CPC) values show how well rules from one year capture rules from the subsequent year. Higher values indicate greater consistency in resistance patterns between consecutive years. (A) Drug-level phenotypic comparisons across all available years for each source. (B) Class-level phenotypic comparisons across all available years. (C) Gene-level genotypic comparisons. (D) Class-level genotypic comparisons. For example, a CPC value of 0.75 indicates that 75 % of the rules from the subsequent year were captured by rules from the previous year.

these comparisons varied by dataset: retail meat phenotypic comparisons span 2002–2021 (excluding 2020), cecal phenotypic comparisons span 2013–2021, and sick cattle phenotypic comparisons span 2018–2022. Genotypic comparisons were limited to more recent years: 2017–2021 for retail meat and cecal data (excluding 2019 for the cecal dataset), and 2018–2022 for sick cattle data. At the drug and gene/mutation level (Fig. 2A,C), phenotypic rulesets are more consistent across time than genotypic rulesets, with cumulative proportion captured (CPC) values generally ranging from 0.4 to 0.8 for phenotypic data compared to 0.2–0.4 for genotypic data. Among phenotypic datasets, cecal and sick cattle phenotypic rulesets show higher temporal consistency than retail meat rulesets, particularly at the drug level (Fig. 2A). When analyzed at the class level (Fig. 2B,D), both phenotypic and genotypic rulesets show improved temporal consistency compared to their respective drug/gene-level analyses (Fig. 2A,C). This improvement is particularly notable for genotypic data, where class-level CPC values approach those of phenotypic data, suggesting that while specific resistance genes may vary year to year, the classes of resistance they confer remain more stable. For example, at the class level, approximately 50 % or more of the MDR patterns in association rules remain consistent between consecutive years across all data sources.

Fig. 3 shows the results of the ruleset comparison between sources.

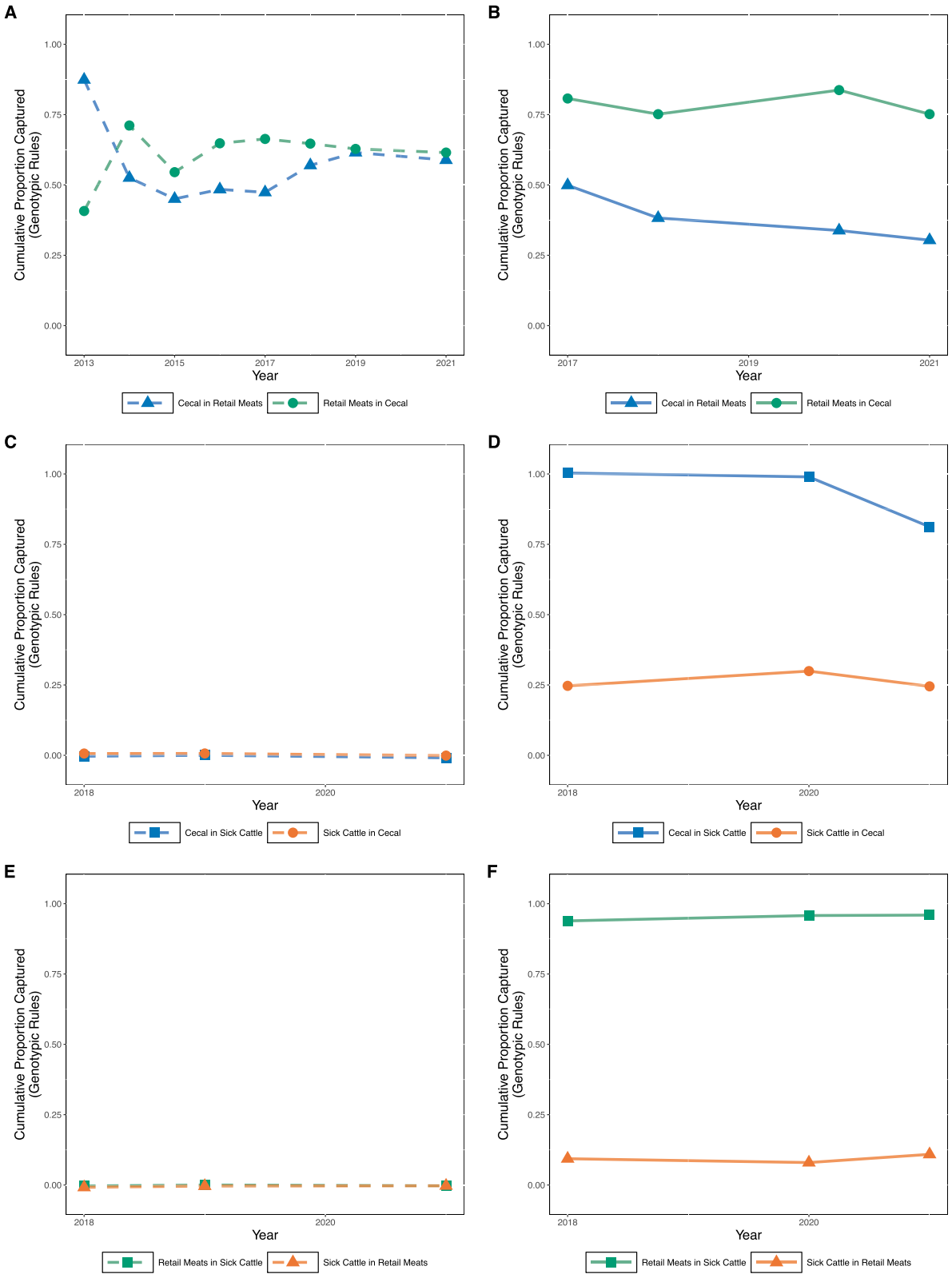


Fig. 3. Comparisons of rulesets between data sources at the antimicrobial class level. Each pair of plots (A,B; C,D; E,F) shows the same reciprocal comparisons using phenotypic data (left plot) and genotypic data (right plot). The cumulative proportion captured (CPC) measures how well rules from one source capture rules from another source, with directionality being important for interpretation. (A,B) Comparisons between cecal and retail meat sources. (C,D) Comparisons between cecal and sick cattle sources. (E,F) Comparisons between retail meat and sick cattle sources. Asymmetry between lines within each panel indicates a directional relationship between sources - for example, if one source's rules better capture another source's rules than vice versa, this suggests the second source may represent a subset of the patterns found in the first.

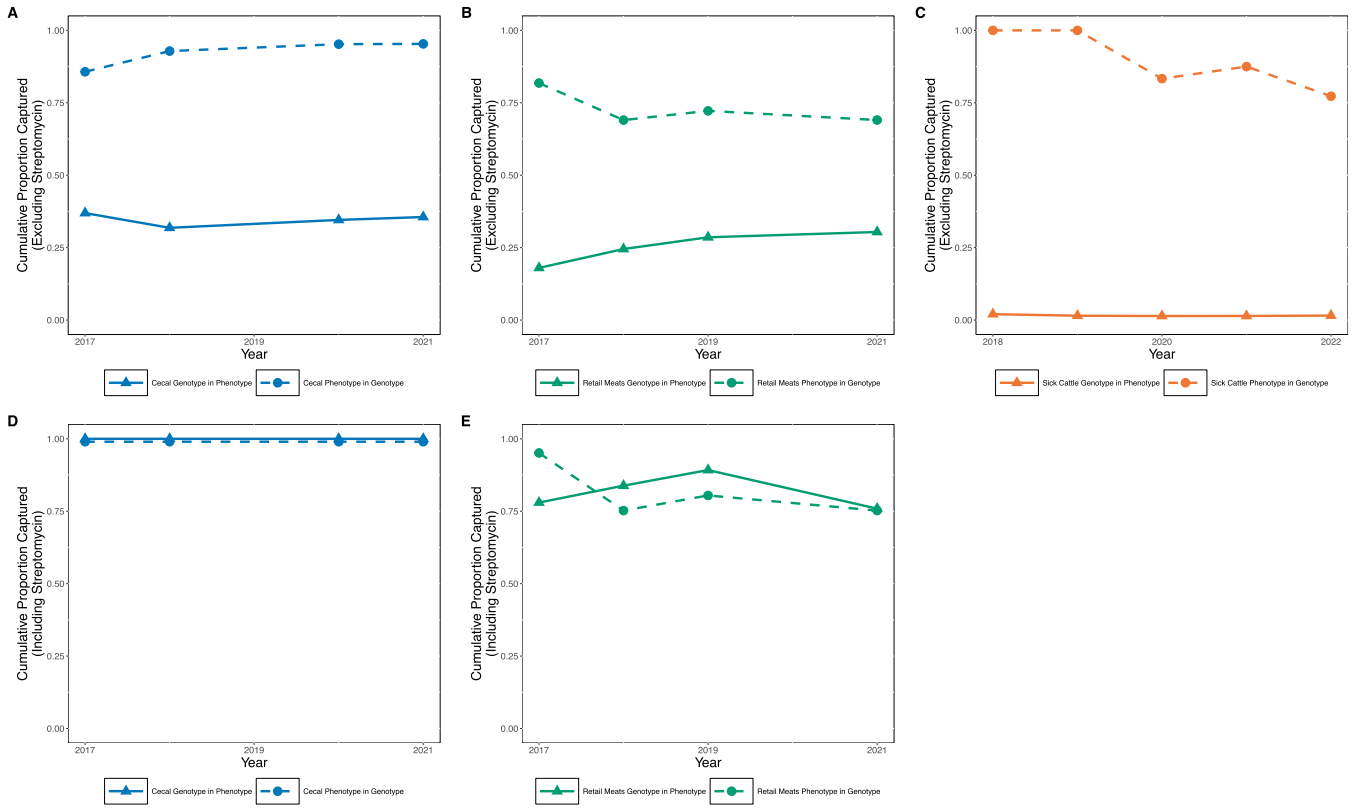


Fig. 4. Comparisons between phenotypic and genotypic rulesets at the antimicrobial class level. Each plot shows reciprocal comparisons between indicators, with dashed lines showing how well genotypic rules capture phenotypic rules and solid lines showing how well phenotypic rules capture genotypic rules. (A,D) Comparisons for cecal data without (A) and with (D) streptomycin included. (C) Comparisons for sick cattle data. (B,E) Comparisons for retail meat data without (B) and with (E) streptomycin included. The asymmetry between solid and dashed lines indicates the degree to which one type of indicator captures the patterns represented by the other. For example, high CPC values for dashed lines combined with low CPC values for solid lines suggest that phenotypic patterns represent a subset of the patterns found in genotypic data.

Fig. 3A, C and E show the comparisons for the phenotypic rulesets and Fig. 3B, D, and F show the comparisons for the genotypic rulesets. For phenotypic comparisons between retail meat and cecal sources, data from 2013 to 2021 were analyzed. Comparisons involving sick cattle phenotypic data were limited to 2018–2022. Genotypic comparisons across all sources were possible only for the period 2018–2021, where data overlapped. All comparisons between sources were made at the class level. For phenotypic rulesets, the cecal dataset was generally better able to capture the retail meat dataset as compared to the reverse. The sick cattle phenotypic rulesets appear to be highly dissimilar to both the retail meat and cecal rulesets; however, this comparison is affected by differences in the set of drugs tested and interpreted for the sick cattle isolates vs retail meat and cecal isolates. Only drugs that were common to both datasets were retained for the comparison, leading to a small

Table 6
Top twenty tabulation patterns recovered by rule mining. Frequency is the number of isolates containing the resistance pattern.

Resistance Pattern	Frequency	Recovered Resistance Pattern as a Rule	Average Cosine	Average Jaccard	Average Kulczynski	Average Support	Source
FIS; TET	415	Yes	0.62	0.39	0.69	0.23	Cecal
FIS; TET	196	Yes	0.6	0.38	0.66	0.11	Retail meat
AMP; CHL; FIS; TET	140	Yes	0.61	0.41	0.65	0.11	Cecal
CHL; FIS; TET	129	Yes	0.63	0.41	0.69	0.15	Cecal
ENR; GEN; NEO; SPE; TIO	122	Yes	0.56	0.37	0.59	0.15	Sick cattle
ENR; NEO; SPE; TIO	103	Yes	0.57	0.38	0.6	0.18	Sick cattle
AMP; TET	86	Yes	0.61	0.38	0.67	0.3	Cecal
NEO; SPE; TIO	71	Yes	0.59	0.4	0.62	0.21	Sick cattle
NEO; SPE	66	Yes	0.59	0.39	0.64	0.25	Sick cattle
AMP; TET	63	Yes	0.55	0.3	0.64	0.18	Retail meat
NEO; TIO	61	Yes	0.58	0.37	0.63	0.25	Sick cattle
ENR; TIO	53	Yes	0.58	0.37	0.63	0.25	Sick cattle
ENR; NEO; TIO	52	Yes	0.57	0.38	0.6	0.21	Sick cattle
ENR; NEO	47	Yes	0.56	0.36	0.61	0.26	Sick cattle
AMP; CHL; COT; FIS; TET	45	Yes	0.6	0.4	0.64	0.1	Cecal
ENR; SPE; TIO	41	Yes	0.58	0.38	0.6	0.2	Sick cattle
AMP; CHL; FIS; TET	38	Yes	0.6	0.4	0.64	0.05	Retail meat
CHL; FIS; TET	37	Yes	0.61	0.39	0.67	0.08	Retail meat
SPE; TIO	37	Yes	0.58	0.38	0.63	0.24	Sick cattle
ENR; NEO; SPE	34	Yes	0.57	0.38	0.6	0.21	Sick cattle

number (n = 2; gentamicin and trimethoprim/sulfamethoxazole) of individual drugs that could be compared and therefore the comparisons for phenotypic data are not informative (Fig. 2 C,E). For the genotypic rulesets, the sick cattle rulesets capture both the retail meat and cecal rulesets exceedingly well, while both the retail meat and cecal rulesets poorly capture the sick cattle rulesets. Additionally, the cecal genotypic rules sets capture the retail meat genotypic rulesets better than the reverse.

Fig. 4 shows the ruleset comparisons between indicators. Originally the analysis excluded streptomycin from the phenotypic data because a change in the interpretive criteria for streptomycin occurred in 2014, which lead to inconsistencies in the data. The genotypic data begins in 2017; so, for purposes of comparison to the genotypic data, streptomycin could be included without producing uninterpretable data. Comparisons between phenotypic and genotypic data were restricted to 2017–2021 for retail meat and cecal data (excluding 2019 for the cecal dataset), and 2018–2022 for sick cattle data. Fig. 2A-C shows the between indicator comparisons when streptomycin was excluded from the phenotypic data. For all sources, the genotypic rulesets capture the phenotypic rulesets exceedingly well, while the phenotypic rulesets poorly capture the genotypic rulesets. Fig. 2D and E shows the between indicator comparisons when streptomycin was included in the phenotypic data; the sick cattle dataset did not include streptomycin. With the inclusion of streptomycin, there is a dramatic improvement in the ability of the phenotypic rulesets to capture the genotypic rulesets.

3.3. Tabulation comparison

The mined rules were compared to resistance patterns generated by tabulation. The top twenty most frequent resistance patterns were inspected. It was found that all the top twenty tabulated resistance patterns were captured by the association rules. In addition to producing the relevant resistance patterns, the association rules provide additional information about the strength of the associations represented by such patterns. This is useful when considering MDR patterns because it provides an estimation of whether the resistances in an MDR pattern are occurring together by chance or whether there is some other mechanism driving the pattern which would be beneficial to further investigate. Table 6 shows the top twenty most frequently occurring resistance patterns.

3.4. Virulence gene and plasmid analysis

Sick cattle isolates (μ=11, σ= 7) showed significantly higher counts of virulence genes compared to both cecal isolates (μ= 6, σ= 4, β = −0.52, 95 % CI [-0.60, −0.45], pMCMC < 0.001) and retail meat isolates (μ= 5, σ= 4, β = −0.70, 95 % CI [-0.78, −0.63], pMCMC < 0.001). Pairwise comparisons revealed significantly higher counts for cecal isolates as compared to retail meat isolates (β = −0.18, 95 % CI [-0.23, −0.13], pMCMC < 0.001).

The cecal data set had an average plasmid-rule mismatch distance of 0.36 and 26.97 % of its matches were perfect. The sick cattle dataset had an average mismatch distance of 0.29 and 32.56 % of its matches were perfect. The retail meat data set had an average mismatch distance of 0.33 and 30.61 % of its matches were perfect. Most mismatches were due to subset relationships. The longest common class combination from the cecal dataset was {aminoglycoside, phenicol, sulfonamide, tetracycline} with an average support of 0.29. The longest common class combination from the cecal dataset was {aminoglycoside, phenicol, sulfonamide, tetracycline} with an average support of 0.29. The sick cattle dataset had three class combination that were of equal length and average support (0.33): {aminoglycoside, phenicol, sulfonamide}, {aminoglycoside, beta-lactam, sulfonamide}, and {aminoglycoside, sulfonamide, tetracycline}. The longest common class combination from the retail meat dataset was {aminoglycoside, sulfonamide, tetracycline} with an average support of 0.06.

4. Discussion

To examine how MDR patterns change across time, rulesets were compared across time holding indicators and sources constant. These comparisons were done at both the drug or gene/mutation level and at the class level. The within indicators comparisons at the class level (see Fig. 2) may be a better reflection of how consistent MDR patterns are across time since MDR is usually defined in terms of resistance to classes rather than to specific drugs or the presence of antimicrobial resistance genes (Magiorakos et al., 2012). The low consistency of the genotypic rulesets across time is likely due to the large number of possible genes, leading to highly variable rulesets. When looking at class level within-indicators comparisons (Fig. 2B,D), the degree of consistency for genotypic and phenotypic rules is more similar as compared to that of the drug and gene/mutation level (Fig. 2A,C). This may suggest that examining how class-level genotypic association rules trend over time can reasonably estimate the degree of consistency in phenotypic MDR patterns. This also suggests that at the class level, on average close to 50 % or more of the MDR patterns remain the same year over year for both genotypic and phenotypic indicators of resistance.

The degree of rule consistency for the cecal and sick cattle phenotypic rulesets is similar whether the analysis is done at the drug level or class level, however the retail meat rulesets are markedly less consistent across time at the drug level than at the class level. This is likely because the retail meat dataset contains several drugs that are inconsistently represented from year to year: there are some drugs that stopped being tested and there are some drugs for which resistance is rare (see Table 2). If the resistance prevalence for a certain drug in a given year is 0 %, then that drug will not show up in the ruleset. If, in another year, the prevalence is non-zero, then these two rulesets will exhibit notable non-overlap. The cumulative proportion captured metric can highlight such cases by virtue of its asymmetry, but the within indicators comparisons were only computed in one direction because it is inherently meaningful to quantify how well a previous year is able to predict a subsequent year. The cecal dataset includes the same drugs as the retail meat dataset; but it spans a smaller range of years, so the fluctuations in rare drug resistances across time have a smaller impact on the rulesets generated from the data. Overall, the difference in the degree of consistency for cecal and retail meat rulesets reflects the difficulty in predicting rare resistances across longer stretches of time.

Comparisons between rulesets generated from different data sources were done to assess how MDR changes across different points in the food production process. Comparisons were made for both phenotypic (Fig. 3A, C, and E) and genotypic (Fig. 3B, D, and F) indicators of resistance at the class level. On the phenotypic side, the sick cattle rulesets were incomparable with the cecal and retail meat data because the sick cattle dataset was tested against a different panel of antimicrobials. For the genotypic ruleset comparisons, the sick cattle rulesets were better able to capture cecal and retail meat rulesets than the reverse. Cecal and retail meat are close to being subsets of the sick cattle genotypic ruleset. For both genotypic and phenotypic rulesets, the cecal rulesets are better able to capture the retail meat rulesets than the reverse. This suggests that as cattle products progress along the food chain (e.g., from farm to cecal to retail meat datasets), only a subset of bacterial isolates move from one point of the food chain to another. Additional isolates may contaminate products at the slaughterhouse, processing, and retail stages, resulting in additional isolate diversity.

This is further supported by the analysis of virulence genes, which provides strong evidence that isolates from clinically ill animals are more likely to be pathogens and are likely more genetically diverse than cecal isolates, and that cecal isolates are more likely to be pathogens and likely more genetically diverse than retail meat isolates. These differences in genetic diversity are reflected in the rulesets. Sick cattle rulesets capturing the other rulesets asymmetrically well, along with the virulence gene analysis, suggests that isolates sampled from sick cattle from farm settings have more diverse patterns of antimicrobial resistance

genes likely driven by increased overall genetic diversity of pathogenic bacteria as compared to commensals. These results also suggest that slaughterhouse settings produce isolates that have more diverse patterns of both antimicrobial resistance genes and phenotypic MDR than those sampled from retail settings. This may be attributable to retail meat having gone through additional processing steps that likely incorporate aseptic efforts as compared to the samples taken at slaughter.

In the comparison of phenotypic and genotypic rulesets, it was found that, when excluding streptomycin, genotypic rulesets captured the phenotypic rulesets well, but the phenotypic rulesets capture the genotypic rulesets poorly (Fig. 4A and B). However, with the inclusion of streptomycin, the phenotypic rulesets captured the genotypic rulesets relatively well (Fig. 4D and E). This suggests that the dramatic asymmetry in the cumulative proportion captured metrics for phenotypic and genotypic rulesets is being driven largely by drug specific aminoglycoside resistance genes in these datasets. Remaining non-overlap in the streptomycin-included analysis may be caused by other drug specific genes for which there is no phenotypic analog in the data (e.g., kanamycin), as well as by genes that are present but fail to confer phenotypic resistance. This may suggest that more drugs need to be tested to fully characterize the variation in MDR patterns.

Given that when enough drugs are assessed, the phenotypic rules and genotypic rules yield a high degree of correspondence, examining the genotypic rules may provide insights into the mechanisms underlying the proliferation of phenotypic MDR patterns. For example, comparing genotypic rules to gene combinations reported to have been found on plasmids suggests that the rules are (at least in part) reflecting the physical structure of the plasmid, with approximately 30 % of plasmid-rule matches being perfect matches. The most common class combinations on plasmids all have common correspondents in their respective phenotypic rulesets, which may suggest the MDR patterns represented by these common phenotypic rules are plasmid driven.

The correspondence between phenotypic and genotypic rules at the class level suggests that longitudinal antimicrobial resistance surveillance programs can use this correspondence to understand long-term trends in resistance and MDR, even when the data type has shifted from phenotypic testing to genomic sequencing (Zankari et al., 2013). This is an important finding because if surveillance programs move towards predominantly genotypic testing, there will still be a means of conducting longitudinal studies.

There are a few limitations to this study. First, we did not include any metadata on the isolates, including geographic origin or history of antimicrobial use because this data is not available for many isolates. Retail meat isolates with a label claim indicating restriction of antimicrobial use have a lower prevalence of MDR than isolates from animals grown with conventional production practices (Stapleton et al., 2024). In addition, MDR in retail meat samples varies according to geographic location and shipping distance (Innes et al., 2023). It is possible that different prevalences of antimicrobial restriction or geographic origin within or across datasets could contribute to dissimilarity among the MDR patterns. Second, genomic information was only available for a subset of isolates, and this limits the analysis of trends of genomic MDR over time and comparisons across datasets. It is unknown if the selection of isolates for genomic sequencing is related or unrelated to their phenotypic characteristics, which could bias the genomic datasets.

Additionally, there were important limitations in the temporal and antimicrobial coverage of our datasets. The temporal coverage of phenotypic and genotypic data was not uniform across sources, with genotypic data generally being available only for more recent years. The antimicrobial panels tested also varied across sources, particularly between NARMS (retail meat and cecal) and NAHLN (sick cattle) datasets. We addressed these limitations through careful data harmonization strategies: temporal comparisons were made within individual datasets to maintain consistency, cross-source comparisons were restricted to overlapping time periods and antimicrobial drugs/classes, and phenotypic-genotypic comparisons were conducted at the class level using only comparable data. While these limitations constrained the scope of certain comparisons, they did not compromise the validity of our findings. The ability to make meaningful comparisons despite such data heterogeneity demonstrates the robustness of our analytical approach. Furthermore, these data limitations reflect real-world challenges in antimicrobial resistance surveillance, where testing methodologies and panels evolve over time and vary across contexts. Understanding how to analyze such heterogeneous data is crucial for leveraging historical surveillance data alongside newer data collection methods.

5. Conclusion

Previous studies have used association rule mining to analyze MDR across time, and here we extend the application of association rule mining to the examination of MDR across multiple sources in the food supply chain and to the analysis of both genotypic and phenotypic data. With this method we were able to pull useful and understandable insights from complex data. More traditional methods of analyzing MDR, such as tabulating the existing MDR patterns, can identify distinct MDR patterns present in a dataset. We show that association rule mining is not only able to identify the same MDR patterns as tabulation methods but is also able to provide insights into the trajectory of those patterns as well as reflect possible mechanistic underpinnings.

Code availability

R code to reproduce this analysis can be found at: https://zenodo.org/records/15085288

CRediT authorship contribution statement

Casey L. Cazer: Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Data curation, Conceptualization. **Laura B. Goodman:** Writing – review & editing, Resources, Conceptualization. **Kristina M. Ceres:** Writing – review & editing, Resources, Conceptualization. **Jordan D. Zehr:** Writing – review & editing, Software, Resources, Methodology. **Gayatri Anil:** Writing – review & editing, Validation, Conceptualization. **Glass Joshua:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization.

Funding

This work was supported by the USDA National Institute of Food and Agriculture, AFRI project 2023–68015–4092.

Declaration of Competing Interest

There are no conflicts of interest associated with this work that could influence our conclusions.

Acknowledgements

The data used in this publication was made possible, in part, by an Agreement from the United States Department of Agriculture's Animal and Plant Health Inspection Service (APHIS). This publication may not necessarily express the views of APHIS.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.prevetmed.2025.106713.

References

Agrawal, R., Imieliński, T., Swami, A., 1993. Mining association rules between sets of items in large databases. Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data. ACM, New York, NY, USA, pp. 207–216. https://doi.org/10.1145/170035.170072.

Brossette, S.E., Sprague, A.P., Jones, W.T., Moser, S.A., 2000. A data mining system for infection control surveillance. Methods Inf. Med. 39, 303–310. https://doi.org/10.1055/s-0038-1634449.

Cazer, C.L., Al-Mamun, M.A., Kaniyamattam, K., Love, W.J., Booth, J.G., Lanzas, C., Gröhn, Y.T., 2019. Shared multidrug resistance patterns in Chicken-Associated escherichia coli identified by association rule mining. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.00687.

Cazer, C.L., Westblade, L.F., Simon, M.S., Magleby, R., Castanheira, M., Booth, J.G., Jenkins, S.G., Gröhn, Y.T., 2021. Analysis of multidrug resistance in staphylococcus aureus with a machine learning-generated antibiogram. Antimicrob. Agents Chemother. 65. https://doi.org/10.1128/AAC.02132-20.

Chang, H.-H., Cohen, T., Grad, Y.H., Hanage, W.P., O'Brien, T.F., Lipsitch, M., 2015. Origin and proliferation of Multiple-Drug resistance in bacterial pathogens. Microbiol. Mol. Biol. Rev. 79, 101–116. https://doi.org/10.1128/mmbr.00039-14.

R. Core Team, 2024. R: A Language and Environment for Statistical Computing.

Dudek, D., 2010. Measures for Comparing Association Rule Sets. pp. 315–322. https://doi.org/10.1007/978-3-642-13208-7_40.

European Committee on Antimicrobial Susceptibility Testing, 2024. MIC and zone distributions and ECOFFs [WWW Document]. ⟨https://www.eucast.org/mic_and_zone_distributions_and_ecoffs⟩. Accessed 11/19/2024.

Feldgarden, M., Brover, V., Haft, D.H., Prasad, A.B., Slotta, D.J., Tolstoy, I., Tyson, G.H., Zhao, S., Hsu, C.-H., McDermott, P.F., Tadesse, D.A., Morales, C., Simmons, M., Tillman, G., Wasilenko, J., Folster, J.P., Klimke, W., 2019. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance Genotype-Phenotype correlations in a collection of isolates. Antimicrob. Agents Chemother. 63. https://doi.org/10.1128/AAC.00483-19.

Feldgarden, M., Brover, V., Gonzalez-Escalona, N., Frye, J.G., Haendiges, J., Haft, D.H., Hoffmann, M., Pettengill, J.B., Prasad, A.B., Tillman, G.E., Tyson, G.H., Klimke, W., 2021. AMRFinderPlus and the reference gene catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci. Rep. 11, 12728. https://doi.org/10.1038/s41598-021-91456-0.

Food and Drug Administration, 2021. NARMS Interpretive Criteria for Susceptibility Testing [WWW Document]. ⟨https://www.fda.gov/media/108180/download⟩.

Food and Drug Administration (FDA), 2024. NARMS Now [WWW Document]. Rockville, MD: U.S. Department of Health and Human Services. Available from URL: https://www.fda.gov/animal-veterinary/national-antimicrobial-resistance-monitoring-system/narms-now-integrated-data. Accessed 12/03/2024.

Giannopoulou, E.G., Kemerlis, V.P., Polemis, M., Papaparaskevas, J., Vatopoulos, A.C., Vazirgiannis, M., 2007. A Large Scale Data Mining Approach to Antibiotic Resistance Surveillance. In Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07). IEEE.

Hadfield, J.D., 2010. MCMC methods for Multi-Response generalized linear mixed models: the MCMCglmm r package. J. Stat. Softw. 33, 1–22.

Hahsler, M., Gruen, B., Hornik, K., 2005. Arules – a computational environment for mining association rules and frequent item sets. J. Stat. Softw. 14, 1–25. https://doi.org/10.18637/jss.v014.i15.

Hahsler, M., Chelluboina, S., Hornik, K., Buchta, C., 2011. The arules R-Package ecosystem: analyzing interesting patterns from large transaction datasets. J. Mach. Learn. Res. 12, 1977–1981.

Hahsler, M., Buchta, C., Gruen, B., Hornik, K., 2023. arules: Mining Association Rules and Frequent Itemsets.

Michael Hahsler, 2015. A Probabilistic Comparison of Commonly Used Interest Measures for Association Rules [WWW Document]. https://mhahsler.github.io/arules/docs/measures.

Harrison, X.A., 2014. Using observation-level randomeffects to model overdispersion in count data in ecology and evolution. PeerJ 2014. https://doi.org/10.7717/peerj.616.

Innes, G.K., Patton, A.N., Nachman, K.E., Casey, J.A., Stapleton, G.S., Abraham, A.G., Price, L.B., Tartof, S.Y., Davis, M.F., 2023. Distance and destination of retail meat alter multidrug resistant contamination in the United States food system. Sci. Rep. 13, 21024. https://doi.org/10.1038/s41598-023-48197-z.

Khachatryan, A.R., Besser, T.E., Hancock, D.D., Call, D.R., 2006. Use of a nonmedicated dietary supplement correlates with increased prevalence of streptomycin-sulfa-tetracycline-resistant escherichia coli on a dairy farm. Appl. Environ. Microbiol. 72, 4583–4588. https://doi.org/10.1128/AEM.02584-05.

Liang, B., Li, X., Zhang, Z., Wu, C., Liu, X., Zheng, Y., 2021. Multidrug resistance analysis method for pathogens of cow mastitis based on weighted-association rule mining and similarity comparison. Comput. Electron. Agric. 190. https://doi.org/10.1016/j.compag.2021.106411.

Love, W.J., Zawack, K.A., Booth, J.G., Gröhn, Y.T., Lanzas, C., 2016. Markov networks of collateral resistance: national antimicrobial resistance monitoring system surveillance results from escherichia coli isolates, 2004-2012. PLoS Comput. Biol. 12. https://doi.org/10.1371/journal.pcbi.1005160.

Love, W.J., Zawack, K.A., Booth, J.G., Gröhn, Y.T., Lanzas, C., 2018. Phenotypical resistance correlation networks for 10 non-typhoidal salmonella subpopulations in an active antimicrobial surveillance programme. Epidemiol. Infect. 146, 991–1002. https://doi.org/10.1017/S0950268818000833.

Ludwig, A., Berthiaume, P., Boerlin, P., Gow, S., Léger, D., Lewis, F.I., 2013. Identifying associations in escherichia coli antimicrobial resistance patterns using additive Bayesian networks. Prev. Vet. Med. 110, 64–75. https://doi.org/10.1016/j.prevetmed.2013.02.005.

Ma, L., Tsui, F.-C., Hogan, W.R., Wagner, M.M., Ma, H., 2003. A framework for infection control surveillance using association rules. AMIA … Annual Symposium Proceedings. AMIA Symposium, pp. 410–414. ⟨http://www.ncbi.nlm.nih.gov/pubmed/14728205⟩. vol. 2003.

Magiorakos, A.P., Srinivasan, A., Carey, R.B., Carmeli, Y., Falagas, M.E., Giske, C.G., Harbarth, S., Hindler, J.F., Kahlmeter, G., Olsson-Liljequist, B., Paterson, D.L., Rice, L.B., Stelling, J., Struelens, M.J., Vatopoulos, A., Weber, J.T., Monnet, D.L., 2012. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clin. Microbiol. Infect. 18, 268–281. https://doi.org/10.1111/j.1469-0691.2011.03570.x.

Marshall, B.M., Levy, S.B., 2011. Food animals and antimicrobials: impacts on human health. Clin. Microbiol. Rev. https://doi.org/10.1128/CMR.00002-11.

Mosaddegh, A., Angel, C.C., Craig, M., Cummings, K.J., Cazer, C.L., 2024. An exploration of descriptive machine learning approaches for antimicrobial resistance: multidrug resistance patterns in salmonella enterica. Prev. Vet. Med. 230, 106261. https://doi.org/10.1016/j.prevetmed.2024.106261.

National Animal Health Laboratory Network, 2023. E. coli from Sick Cattle: Private Data Sharing Agreement [dataset].

Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: convergence diagnosis and output analysis for MCMC. R. N. 6, 7–11.

Reference Gene Catalog [WWW Document], n.d. https://www.ncbi.nlm.nih.gov/pathogens/refgene/.

Stapleton, G.S., Innes, G.K., Nachman, K.E., Casey, J.A., Patton, A.N., Price, L.B., Tartof, S.Y., Davis, M.F., 2024. Assessing the difference in contamination of retail meat with multidrug-resistant bacteria using for-consumer package label claims that indicate on-farm antibiotic use practices— United States, 2016–2019. J. Expo. Sci. Environ. Epidemiol. 34, 917–926. https://doi.org/10.1038/s41370-024-00649-y.

Stokes, H.W., Gillings, M.R., 2011. Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. FEMS Microbiol. Rev. https://doi.org/10.1111/j.1574-6976.2011.00273.x.

Tan, P.-N., Kumar, V., Srivastava, J., 2002. Selecting the right interestingness measure for association patterns. Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, USA, pp. 32–41. https://doi.org/10.1145/775047.775053.

Tange, O., 2015. GNU Parallel 20150322 ('Hellwig'). Zenodo.

The National Antimicrobial Resistance Monitoring System Strategic Plan 2021-2025 [WWW Document], 2021. ⟨https://www.fda.gov/media/79976/download?attachment⟩.

Tian, R., Zhou, J., Imanian, B., 2024. PlasmidHunter: accurate and fast prediction of plasmid sequences using gene content profile and machine learning. Brief. Bioinform. 25. https://doi.org/10.1093/bib/bbae322.

USDA APHIS V.S. National Animal Health Laboratory Network (NAHLN) Antimicrobial Resistance Pilot Project, 2020.

Wagner, B.A., Dargatz, D.A., Morley, P.S., Keefe, T.J., Salman, M.D., 2003. Analysis methods for evaluating bacterial antimicrobial resistance outcomes. Am. J. Vet. Res. 64, 1570–1579.

Zaharia, C., Muresan, R., Moleriu, R., Zaharie, D., 2019. Analysis of association measures used to discover antimicrobial resistance patterns. 2019 E-Health and Bioengineering Conference (EHB), pp. 1–4. https://doi.org/10.1109/EHB47216.2019.8969945.

Zaharia, C., Muresan, R., Zaharie, D., 2021. Can association rules be used to improve additive Bayesian network models? 2021 9th E-Health and Bioengineering Conference, EHB 2021 https://doi.org/10.1109/EHB52898.2021.9657656.

Zankari, E., Hasman, H., Kaas, R.S., Seyfarth, A.M., Agerso, Y., Lund, O., Larsen, M.V., Aarestrup, F.M., 2013. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. J. Antimicrob. Chemother. 68, 771–777. https://doi.org/10.1093/jac/dks496.

Zawack, K., Li, M., Booth, J.G., Love, W., Lanzas, C., Gröhn, Y.T., 2016. Monitoring antimicrobial resistance in the food supply chain and its implications for FDA policy initiatives. Antimicrob. Agents Chemother. 60, 5302–5311. https://doi.org/10.1128/AAC.00688-16.

Zawack, K., Love, W., Lanzas, C., Booth, J.G., Gröhn, Y.T., 2018. Inferring the interaction structure of resistance to antimicrobials. Prev. Vet. Med. 152, 81–88. https://doi.org/10.1016/j.prevetmed.2018.02.007.