Applying Association Rule Mining to the Analysis of Multidrug Resistance Patterns Using Phenotypic and Genotypic Indicators

Joshua Glass[1], Gayatri Anil[1,2], Kristina M. Ceres[2], Laura B. Goodman[2], Casey L. Cazer[1,2]

[1]Department of Clinical Sciences, College of Veterinary Medicine, Cornell University

[2]Department of Public and Ecosystem Health, College of Veterinary Medicine, Cornell University

Background

Antimicrobial Resistance

- Acquired resistance generally refers to the nonsusceptibility of an organism to an antimicrobial against which the organism in question is not intrinsically resistant.
- Multi-drug resistance (MDR) refers to cases where organisms have acquired resistance to antimicrobials from 3 or more antimicrobial classes[1].
- MDR is a public health threat because it reduces treatment options for animal and human illness.
- Agricultural antimicrobial use can select for resistance.
- This study examines MDR at different points in the food supply chain.
- The analysis of MDR is difficult due to the quantity and complexity of possible MDR patterns.

Association Rule Mining

- Association rule mining has historically been used to analyze consumer purchasing behavior.
- The Apriori algorithm allows for the efficient analysis of complex associations in large, computationally expensive datasets[2,3].
- Previous work has shown that association rule mining is suitable for the analysis of MDR[4,5].
- The current work uses association rule mining to analyze MDR in cattle-associated *Escherichia coli*.
- The R package arules was used to implement the Apriori algorithm[6,7].
- The current work incorporates a set of data-driven approaches for the selection of various modeling parameters and analysis decisions.

Methods

Data

- Data were retrieved from the National Antimicrobial Monitoring System (NARMS).
- Data were from two sources: retail meat samples and cecal (slaughterhouse) samples.
- Retail meat data included 4586 isolates from 2002 to 2019.
- Cecal data included 8346 isolates from 2013 to 2021.
- Data consisted of minimum inhibitory concentrations (MIC) for 12 to 15 common antimicrobials, as well as genotypic data staring in 2017.

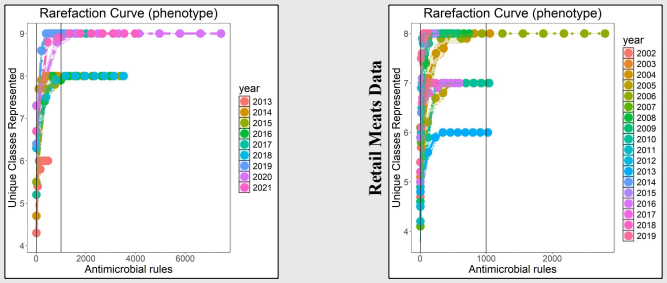
Dichotomization

- Each MIC was dichotomized as either susceptible or resistant using NARMs interpretative criteria.
- Genotype data was dichotomized as presence or absence of each resistance gene in the dataset.

1 Measure Selection

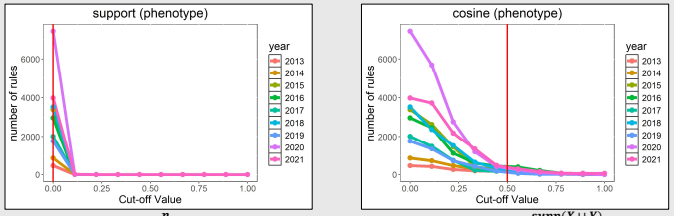
Quality measures were selected based on PCA[4] and on desired properties.
In addition to support, cosine, jaccard, and Kulczynski were selected because they are symmetrical, null-invariant[8,9] and were represented in the PCA top loadings.

2 Target Rule Quantity Selection



Plots showing minimum number of rules needed to represent all antimicrobial classes in the dataset.

3 Cut-Off Selection

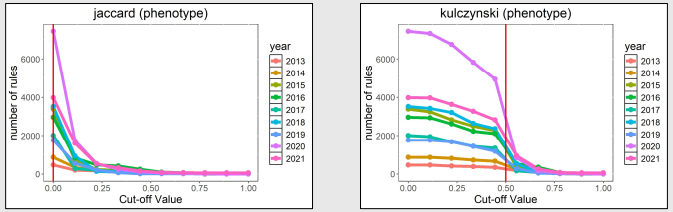


$$supp(X) = \frac{n_x}{n} = P(X)$$

Range: [0, 1]
Selected Cut-Off: 0.0

$$cosine(X \rightarrow Y) = \frac{supp(X \cup Y)}{\sqrt{(supp(X)supp(Y))}}$$

Range: [0, 1] (0.5 means no correlation)
Selected Cut-Off: 0.5



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

Range: [0, 1]
Selected Cut-Off: 0.0

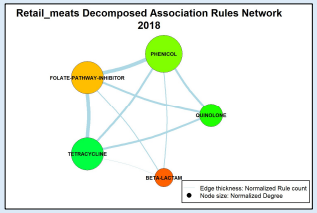
$$kulc(X \rightarrow Y) = \frac{1}{2}(P(X|Y) + P(Y|X))$$

Range: [0, 1] (0.5 typically means uninteresting)
Selected Cut-Off: 0.5

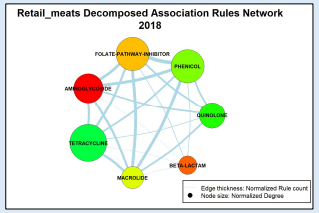
Results

Comparing Rule-Sets Graphically

Phenotype



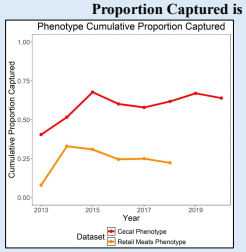
Genotype

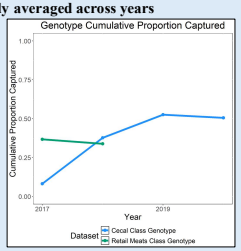


Comparing Rule-Sets Numerically

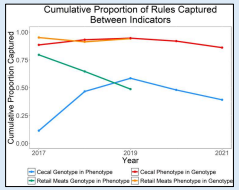
$$Proportion\ Captured = \frac{|R_1 \cap R_2|}{|R_1|}$$ R_1: Rule Set 1; R_2: Rule Set 2;

Proportion Captured is progressively averaged across years

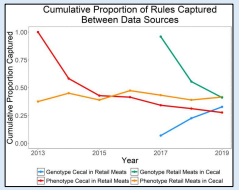




R_1 = rule set from year i; R_2 = rule set from year i+1.



R_1 = rule set from resistance indicator 1;
R_2 = rule set from resistance indicator 2.



R_1 = rule set from data source 1;
R_2 = rule set from data source 2.

Conclusions

- Association rule mining was able to effectively characterize MDR patterns and the associations among them.
- Rule set comparisons showed that, in both data sources, the genotypic data contained rules unaccounted for by the phenotypic data, whereas genotypic data captured the phenotypic rules well. This discrepancy seems to be driven by drug-specific aminoglycoside resistance genes as exemplified by the network graphs presented. This may mean that more drugs should be tested phenotypically to have a more complete picture of MDR patterns and associations.
- Rule set comparisons also, generally, showed greater rule consistency in the cecal data. Additionally, the cecal more consistently and better represented retail meats rules as compared to how well the retail meats data represented the cecal rules. This greater consistency and cross-data source generalizability may suggest that the cecal data better represents MDR in cattle associated *E. coli*.

Acknowledgements

This work is supported by the USDA National Institute of Food and Agriculture, AFRI project 2023-68015-4092

Link to References and Supplemental Materials:

Contact: jg2527@cornell.edu