Supplementary Appendix S1

Bacteriology

Samples for aerobic culture were directly inoculated onto trypticase soy agar with 5% sheep blood (BD BBL) and Chocolate II Agar (BD BBL) for non-selective growth, and on selective agars for the isolation of Gram negative (Eosin Methylene Blue agar; BD BBL) and Gram positive (Columbia CNA agar with 5% sheep blood; BD BBL)) organisms. Samples were also enriched with Chopped Meat Medium (CMC) with Carbohydrates (Anaerobe Systems). Occasionally, joint fluid samples were received inoculated into enrichment broth bottles typically used for blood cultures. In these cases, the broth was inspected for turbidity on receipt and if it appeared turbid, it was sub-cultured to agar plates immediately. If it appeared non-turbid, the bottle was aseptically vented and incubated in ambient air at 35±2°C for 18-24 hours, and then sub-cultured to selective and non-selective plates as previously described.

Agar plates were incubated in 6±1% CO2 at 35±2°C, while the CMC broth was incubated anaerobically at 35±2°C. All were evaluated for bacterial growth at 24- and 48-hours incubation. If the CMC broth showed turbidity during the incubation period, it was sub-cultured to trypticase soy agar with 5% sheep blood, Chocolate II and Brucella agar (Anaerobe Systems). Blood and Chocolate agars were incubated in 6±1% CO2 at 35±2°C, while the Brucella agar was incubated anaerobically at 35±2°C, for 24 and 48 hours.

Identification of bacteria was performed either using the Sensititre automated bacterial identification system (GNID/GPID, Thermo Scientific) in combination with 16S rRNA sequencing

PCR as needed, or by matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF). A Bruker Microflex LT/SH instrument was used for MALDI-TOF, employing the Real Time Classification software (RTC) version 3.1 or the MBT Compass 4.1 software.

Susceptibility testing was performed using the Sensititre system according to the recommendations of the manufacturer. Briefly, colonies taken from a non-selective agar plate were inoculated into sterile water to achieve a MacFarland Standard of 0.5. Then, depending on the organism, either 1, 10, 30 or 50 µl of this suspension was inoculated into 11 ml of cation-adjusted Mueller Hinton-broth with N-Tris(hydroxymethyl) methyl-2-aminoethane sulfonic acid (CaMHB w/TES) or CaMHB w/TES and lysed horse blood. Antimicrobial susceptibility test plates provided by Sensititre were then inoculated with this medium with 50 µl in each well. Plates were incubated at 35±2°C without CO2 for 18-24 hours (depending on organism) and examined for growth by increasing turbidity in the wells to determine the minimum inhibitory concentration (MIC). Interpretations of MICs were performed in accordance with guidelines established by the Clinical Laboratory Standards Institute (CLSI).

Multidrug Resistance Analysis

A machine learning approach is employed to investigate correlations between multiple antimicrobial resistances for gram-positive and gram-negative bacteria. Since association rules demonstrated their great potential to obtain hidden co-occurrence relationships within transactional databases, they have been increasingly applied in different fields [1] including

analysis of multidrug resistance [2, 3]. Association rules are the statements that represent the relationship between items in data transactions [5]. Let I = {I1, I2, …, In} be a set of items and D = {T1, T2, …, Tm} be a set of data transactions where each transaction T contains a subset of items in I. Let X and Y be a set of items (itemset) where X, Y ⊆ I. In the application to MDR, each bacterial isolate is a transaction and the antimicrobial susceptibility tests, and other co-variates such as year or gram classification, are the items.

An association rule is represented in the form X→Y, where X called the antecedent, and Y called consequent, where (X∩Y=φ) and implies that the antecedent itemset (X) have a co-occurrence relation with the consequent itemset (Y) in transactions existing in D. Therefore, association rules can be used as a method for extracting hidden relationships among items within transactional databases. In the context of MDR, the association rules represent relationships between different antimicrobial resistances and other co-variates.

Typically, three measure are used to determine the relevance of association rules. These measures, called interestingness measures, include support, confidence, and lift, which are defined as follows:

- Support (supp) of an association rule X → Y is the proportion of transactions containing both itemsets X and Y out of the total number of transactions in D. It is the probability of X and Y, $P(X \cup Y)$,the same as the prevalence of X and Y in the dataset.

- Confidence (conf) of an association rule X → Y represents the proportion of transactions containing itemset X which also contains Y. This is the same as the conditional probability of Y given X, $\frac{P(X \cup Y)}{P(X)}$.
- Lift (lift) of an association rule X → Y represents the increase in probability of occurrence of Y because of presence of X. It is calculated as $\frac{conf(\mathrm{X} \rightarrow \mathrm{Y})}{P(Y)} = \frac{P(X \cup Y)}{P(X)P(Y)}$. A value of 1 indicates that the appearance of the consequent and the antecedent in the rule is independent. On the other hand, lift values greater than 1 indicate a meaningful correlation between the antecedent and the consequent.

The FP-Growth [4] is one of most common algorithms for association mining, especially for big-data analysis. We used a Parallel FP-Growth [5] on PySpark, a Python library over the SPARK platform, besides two-step Robust Association Rules (RAR) approach [6] to deal with incomplete item-sets. RAR disregards the null values in item-sets and applies the support and confidence thresholds on non-null values. This is important for MDR analysis when the bacterial isolates are not all tested consistently against the same antimicrobials.

For this analysis, we extracted rules that met 0.05 support, 0.1 RAR support, 0.8 confidence and 1.5 lift thresholds on antimicrobials (items) that are commonly used for treating infections in horses: gentamicin, chloramphenicol, penicillin, trimethoprim-sulfamethoxazole, tetracycline, sulphadimethoxine, enrofloxacin, and amikacin.

First, associations rules of gram-positive and gram-negative were extracted separately and associations between antimicrobials were visualized using Chord diagrams, which presents associations between items (antimicrobial resistance) using arcs proportional to the importance of associations. Each association rule was divided into antecedent-consequent pairs. If more than one antimicrobial resistance was in the antecedent, each was paired with the consequent. For example, a rule [X, Z] → Y would yield two antecedent-consequent pairs: X → Y and Z → Y. For each antecedent-consequent pair, lift values were summed across all association rules containing that pair. Next, the summed lift values were scaled and used as weights to determine the number of arcs in the Chord diagram; more arcs indicate a higher total lift between two items. Association rules that met the support, confidence, and lift thresholds in at least one study period were extracted and compared by changes in support, confidence, and lift factors between study periods for gram-negative and gram-positive isolates. Changes by 25% or more in support or lift, and 10% or more changes in confidence were considered major differences.