DAYZERO DIAGNOSTICS

# Quantification of Predictive Power of Genomic Resistance Locus Databases Reveals Potential Limitations of Marker-Based Diagnostics

Miriam H. Huntley, PhD, Day Zero Diagnostics, Boston, Massachusetts

## Background

The growth of antimicrobial resistance is emerging as a public health crisis, hastening the need for rapid diagnostics that can measure the antibiotic resistance profile of a clinical infection and allow healthcare providers to quickly determine targeted, effective treatments for patients. Many of the emerging rapid diagnostics for antibiotic resistance rely on marker-based technology, for example PCR assays, which detect the presence or absence of a panel of resistance markers. The sensitivity of these diagnostics is bounded by the amount of clinical resistance that can in fact be predicted by known resistance loci. Here we analyze the clinical antibiotic resistance accounted for by three genomic resistance locus databases (CARD, ARG-ANNOT, and ResFinder) which list resistance conferring genes and mutations, and in some cases the antibiotic drug or drug category that are affected by the presence of these loci. Though these resistance databases do not claim to account for all published resistance loci in the literature, they represent the most comprehensive curated collections to date.

## Approach

We measure the predictive power of these databases in three different publicly available genotype-phenotype datasets, consisting of whole genome sequencing data of Staphylococcus aureus, Streptococcus pneumonia, and Mycobacterium tuberculosis clinical isolates, matched with their antibiotic resistance phenotypic measurements. Presence of resistance loci (genes and mutations) is determined by performing a BLAST search and gene reconstruction from assembled sequencing data. For each putative resistance locus, we measure the area under the curve (a.u.c.), sensitivity, and specificity that result from using the locus to predict resistance.

## Results

Aggregating resistance prediction across all loci which pass a quality filter, we find that the prediction accuracy of these databases can differ dramatically between different species and drugs, for example attaining 0.98 a.u.c. for methicillin resistance prediction in S. aureus, while only achieving 0.36 a.u.c. for trimethoprim resistance prediction in S. pneumonia.

## Conclusions

We conclude that while prediction sensitivity of genomic resistance databases will vary with species and drug, there is still significant resistance not accounted for by these databases. Whether this resistance is mediated by as yet undiscovered genomic mechanisms or alternatively by non-genomic mechanisms (e.g. through transcription) remains to be explored.

 