



# Keynome G-AST: Development of A Novel Machine Learning Method for Determining Bacterial Antibiotic Susceptibility from Genomic Sequences

J. Wittenbach<sup>1</sup>, P. Knysh<sup>1</sup>, P. Ramaraj<sup>1</sup>, D. Surujon<sup>1</sup>, T. Farrell<sup>1</sup>, I. Herriott<sup>1</sup>, M. Sater<sup>1</sup>

<sup>1</sup> Day Zero Diagnostics, Boston, Massachusetts

## Background

The current gold-standard approach for bacterial antibiotic susceptibility testing (AST) is a culture-based method, which suffers from a slow multi-day turnaround time and fails entirely when the pathogen does not grow. Genomic sequencing has emerged as a method for determining the AST profile of bacterial strains, with “lookup” algorithms that test for the presence/absence of known genomic resistance loci (e.g. *mecA* gene in *Staphylococcus* methicillin resistance). These approaches work well in instances of simple resistance but have low accuracy in cases of complex resistance. Here we present Keynome g-AST (*genomic*-AST), a data-driven approach to determining AST from genomic sequences that leverages a large-scale dataset and machine learning methods to accurately predict resistance.

## Methods

We built MicrohmDB, a large-scale dataset consisting of the whole genome sequences of over 45,000 microbial pathogens matched with their phenotypically derived AST results. Samples spanning over 100 bacterial species were collected over a 5-year period from multiple geographic sites, and sequencing and phenotypic data underwent rigorous quality control. We trained and validated Keynome g-AST kmer-based machine learning models on this MicrohmDB dataset to predict AST from genomic sequences and measured accuracy against the phenotypic AST results. Models predicted susceptibility/non-susceptibility and Minimum Inhibitory Concentration, and accuracies were assessed on both shortread (Illumina) and long-read (Oxford Nanopore Technologies - ONT) sequencing data.

## Results

Keynome g-AST achieved high accuracy on a broad range of gram positive/negative species spanning multiple drug classes. Susceptibility/non-susceptibility predictions showed >95% positive and negative percent agreement on over 60 species/drug combinations. The models achieved a high degree of concordance (>97-99%) between Illumina and ONT sequencing data. Keynome g-AST also showed marked improvement for many species/drug combinations when compared with a presence/absence “lookup” method.

## Conclusion

Keynome g-AST presents a promising method for determining the AST profile of pathogens directly from genomic sequencing data. When combined with sample preparation methods for sequencing pathogens directly from clinical samples, and rapid sequencing technologies such as ONT sequencing, Keynome g-AST can enable rapid culture-free AST diagnostics.