NEWS & ANALYSIS

GENOME WATCH

Breaking the code of antibiotic resistance

Stephanie W. Lo, Narender Kumar and Nicole E. Wheeler

This month's Genome Watch highlights how a better understanding of genotype–phenotype correlation may lead to the design of new diagnostic tests for antimicrobial resistance in clinical settings.

Whole-genome sequencing (WGS) offers the potential to predict antimicrobial susceptibility from a single assay in health-care settings and expedite the investigation of resistance in slow-growing bacteria such as *Mycobacterium tuberculosis*. Despite advantages over traditional antimicrobial susceptibility testing (AST) of bacteria, WGS-inferred AST has yet to be applied to guide clinical decisions[1]. One reason for this is our incomplete knowledge of how genetic variants correlate with susceptibility for specific antimicrobials. To address this, recent studies have used genome-wide association studies (GWAS) and machine learning to identify previously unknown resistance determinants and assess the effect of SNPs or genes on resistance in various bacterial species.

In one of the largest and most comprehensive studies published to date, Coll et al. used the combined power of WGS and GWAS to uncover novel mutations associated with resistance to cycloserine, ethionamide and para-aminosalicylic acid from a collection of 6,450 *M. tuberculosis* isolates[2]. The analysis highlighted the importance of including small indels and large deletions for improving the predictability of resistance phenotypes. The authors also uncovered a number of epistatic interactions, shedding light on the compensatory mechanisms employed by the bacterium. Studies such as this have advanced our understanding of the complex mechanisms governing bacterial phenotypes and expanded the evidence base for the development of molecular diagnostic tests in clinical settings. However, the analysis output is yet to be translated into clinical settings for decision making.

Ideally, the bioinformatics pipeline for clinical use should be designed to yield quantitative measures of resistance, such as minimum inhibitory concentration (MIC), which can be used to guide clinical decisions. Li et al. developed a machine-learning model to predict MICs of six β-lactam antibiotics based on the amino acid sequences of three penicillin-binding proteins (PBPs) in *Streptococcus pneumoniae*[3]. The model was trained with a data set of 2,528 invasive isolates and then challenged by another collection of 1,781 isolates that contained 109 new PBP sequences that the model had not encountered during training. The MIC predictions showed >97% agreement with phenotypic MICs within ±1 dilution. The very major discrepancy (resistant isolate predicted as sensitive) rate was 1.4%. Incorporating newly identified mutations in the model may improve the accuracy of predictions.

Nguyen et al. also tested the possibility of using machine-learning to predict the MICs of a comprehensive panel of 20 antibiotics for *Klebsiella pneumoniae*[4]. They used overlapping decanucleotides from whole-genome assemblies as input that requires no a priori knowledge of the encoded gene content. The overall accuracy of their model was 92%, ranging from 61–100% for individual antibiotics, with accuracy being largely dependent on the number of resistant isolates that were sampled for each antibiotic. The authors also compared the model that used entire genomes with alternatives that used only known antibiotic resistance genes or non-antibiotic resistance genes. Surprisingly, the overall accuracy of these models remained at 92% and the accuracy for each antibiotic was also nearly identical, suggesting that known antibiotic resistance genes are sufficient for performing MIC prediction in *K. pneumoniae*, whose antibiotic resistance is usually conferred by acquisition of antibiotic resistance genes. The comparison also revealed that the analysis of non-antibiotic resistance genes, possibly the resistance-associated mobile genetic elements (for example, *ISEcp1* associated with bla_{CTX-M}), could predict resistance without any information from the causative genes or mutations in *K. pneumoniae*. This study provided a comprehensive framework for building MIC prediction models for other pathogenic bacteria.

In summary, these recent studies demonstrate the value in combining the insights we gain into the genetic determinants of resistance from GWAS with machine-learning to accurately predict resistance. The competitive performance of these methods indicates that they may enable us to make more accurate clinical predictions using raw sequencing data.

Stephanie W. Lo and Nicole E. Wheeler are at the Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK. Narender Kumar is at the Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK; and the Department of Medicine, University of Cambridge, Cambridge, UK.

microbes@sanger.ac.uk

1. Ellington, M. J. et al. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. *Clin. Microbiol. Infect.* **23**, 2–22 (2017).
2. Coll, F. et al. Genome-wide analysis of multi- and extensively drug-resistant *Mycobacterium tuberculosis*. *Nat. Genet.* **50**, 307–316 (2018).
3. Li, Y. et al. Validation of beta-lactam minimum inhibitory concentration predictions for pneumococcal isolates with newly encountered penicillin binding protein (PBP) sequences. *BMC Genomics* **18**, 621 (2017).
4. Nguyen, M. et al. Developing an in silico minimum inhibitory concentration panel test for *Klebsiella pneumoniae*. *Sci. Rep.* **8**, 421 (2018).

Competing interests
The authors declare no competing interests.

Philip Patenall/Macmillan Publishers Limited