



# Genome-Based Prediction of Bacterial Antibiotic Resistance

Michelle Su,<sup>a,b,c</sup> Sarah W. Satola,<sup>a,c,d</sup> Timothy D. Read<sup>a,c,d</sup>

<sup>a</sup>Department of Infectious Diseases, Emory University, Atlanta, Georgia, USA

<sup>b</sup>Antimicrobial Resistance and Therapeutic Discovery Training Program, Emory University, Atlanta, Georgia, USA

<sup>c</sup>Antibiotic Resistance Center, Emory University, Atlanta, Georgia, USA

<sup>d</sup>Emory Investigational Clinical Microbiology Laboratory, Emory University, Atlanta, Georgia, USA

**ABSTRACT** Clinical microbiology has long relied on growing bacteria in culture to determine antimicrobial susceptibility profiles, but the use of whole-genome sequencing for antibiotic susceptibility testing (WGS-AST) is now a powerful alternative. This review discusses the technologies that made this possible and presents results from recent studies to predict resistance based on genome sequences. We examine differences between calling antibiotic resistance profiles by the simple presence or absence of previously known genes and single-nucleotide polymorphisms (SNPs) against approaches that deploy machine learning and statistical models. Often, the limitations to genome-based prediction arise from limitations of accuracy of culture-based AST in addition to an incomplete knowledge of the genetic basis of resistance. However, we need to maintain phenotypic testing even as genome-based prediction becomes more widespread to ensure that the results do not diverge over time. We argue that standardization of WGS-AST by challenge with consistently phenotyped strain sets of defined genetic diversity is necessary to compare the efficacy of methods of prediction of antibiotic resistance based on genome sequences.

**KEYWORDS** antibiotic resistance, genome-based prediction

## IMPORTANCE OF AST

Antibiotic-resistant bacterial infections are a global threat. Many previously manageable bacterial infections are becoming increasingly hard to treat. The CDC recently estimated that in the United States alone, at least two million people will be infected by a drug-resistant bacterium each year, and at least 23,000 people will die as a result (1). Rising rates of resistance amplify the morbidity and economic burden associated with infections. Even successful treatment can come with increased complications, as more-toxic drugs of last resort, like colistin, are being used more frequently because bacteria are not susceptible to less-toxic antibiotics. For the management of infections in both the clinic and community, accurate detection of antimicrobial resistance is necessary to guide treatment decisions.

Culture-based antimicrobial susceptibility testing (AST) is still the primary method employed by clinical laboratories. While there are other promising approaches for phenotypic detection and rapid nonsequencing genetic methods currently in use (e.g., PCR for resistance determinants) (2), dramatic progress over the past 5 years in the applications of genomics has caught the attention of the clinical microbiology community. Whole-genome sequencing for antimicrobial susceptibility testing (WGS-AST) offers the potential to provide rapid, consistent, and accurate predictions of every known resistance phenotype for a strain, as well as simultaneously provide rich surveillance data. Recent reviews of the subject have focused on clinical standardization (2–4). Here, we concentrate on the problem of predicting resistance based only on genome sequence and consider the future symbiotic relationship between genomics and phenotypic-based AST.

**Citation** Su M, Satola SW, Read TD. 2019. Genome-based prediction of bacterial antibiotic resistance. *J Clin Microbiol* 57:e01405-18. <https://doi.org/10.1128/JCM.01405-18>.

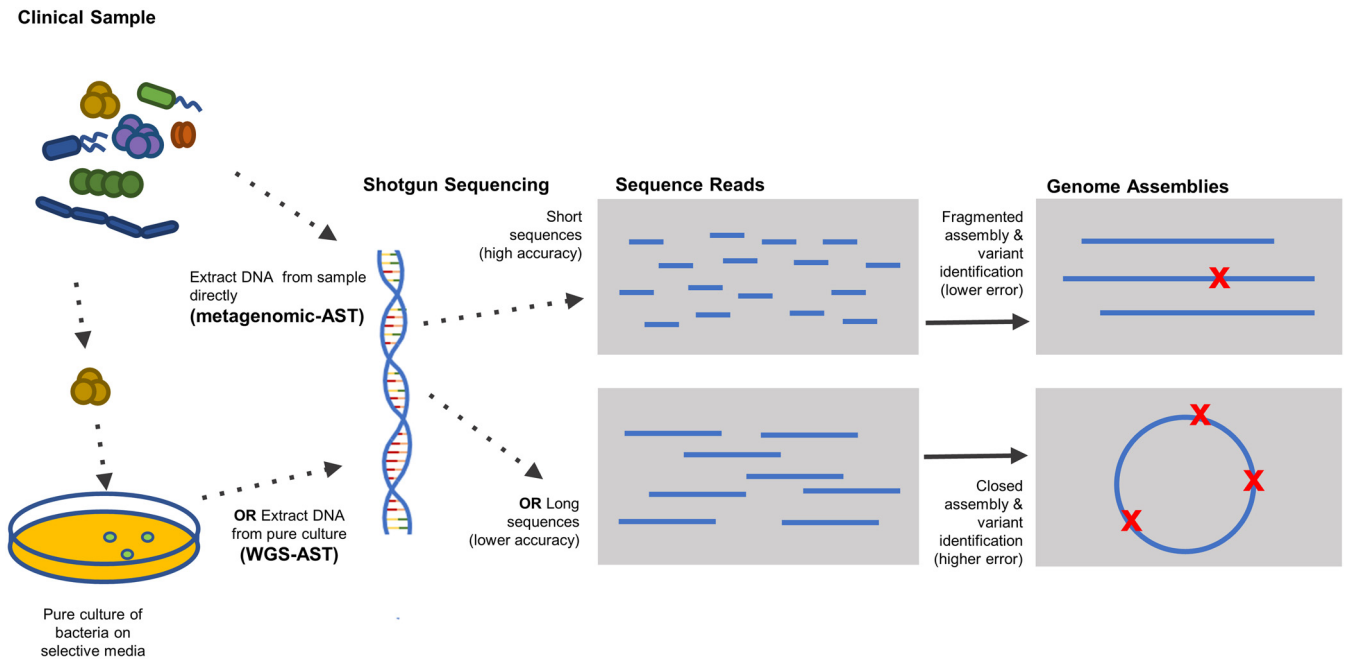
**Editor** Alexander J. McAdam, Boston Children's Hospital

**Copyright** © 2019 Su et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Timothy D. Read, [tread@emory.edu](mailto:tread@emory.edu).

**Accepted manuscript posted online** 31 October 2018

**Published** 27 February 2019



**FIG 1** Overview of genome sequencing and how it is used in WGS-AST. DNA is extracted from directly from bacteria in clinical samples (metagenomics) or more commonly, from cultured bacterial colonies. Sequencing technologies fragment the DNA and then randomly sequence to produce a library of reads (stored in FASTQ files). The reads are assembled into genomic scaffolds *in silico*. Sequencing is performed either using short-read second-generation technology, which tends to produce fragmented whole-genome assemblies of high accuracy, or long-read third-generation technologies that have higher error rates but more complete assemblies. WGS-AST algorithms operate on the raw reads and/or assembled contigs.

### POTENTIAL ADVANTAGES OF WGS-AST

WGS-AST follows selective culturing of the bacterium of interest from a clinical sample (Fig. 1). AST following direct shotgun sequencing of clinical samples (metagenomic-AST) is also possible (see, e.g., reference 5) and is the subject of intense current research, but it is more complex, expensive, and prone to false-negative results due to the potentially low abundance of the pathogen of interest relative to host DNA. Metagenomic-AST also includes approaches that enrich a library of antibiotic resistance DNA fragments from complex clinical samples before sequencing (6). Slow-growing or hard-to-culture bacteria (such as *Mycobacterium tuberculosis* [6, 7]) are important early targets for metagenomic-AST because DNA sequencing may be easier and faster than obtaining enough culture growth for phenotypic testing. This review will focus on the WGS-AST application, but many principles may also apply to other sequence-based AST approaches. Compared to culture-based AST or nucleic acid amplification tests (NAATs), which are often limited by the number of resistant phenotypes that can be determined from one test, WGS-AST can ascertain the antibiotic resistance phenotypes of the entire genome simultaneously, and phenotypes where multiple loci contribute can be easily screened (instead of performing multiplex PCRs). Once collected, the genome sequence data are stored digitally and can be queried for other purposes (e.g., complete-genome multilocus sequence typing [cgMLST] genotype [8] and virulence [9]). Genomes can be sequenced to very high levels of depth, giving very accurate sequence data. Unlike NAATs, there is no reliance on primer specificity for template amplification, reducing the possibility of false-negative results. The accumulation of genomes in clinical laboratories creates a data source that can be used to survey the evolution of pathogens (10). If new antibiotic resistance loci are discovered, these databases can be immediately scanned to understand how long these genes have been circulating and how they may have entered the clinical setting. One of the first examples of this type of retrospective surveillance investigation was performed for the emergence of *mcr-1* colistin resistance in Germany (11, 12), where a database of 577 *Enterobacteriaceae*

**TABLE 1** Genomics terms

Term	Definition <sup>a</sup>
Quality score	A measure of the probability of an inaccurate base call, typically represented by the Phred score [ $Q = -10 \log_{10}(P)$ ] (113). A Q of 10 represents a 1 in 10 chance of error, whereas a Q of 40 is a 1 in 10,000 chance of error.
Coverage	A measure of how many instances a base was sequenced, as quantified by number of unique reads mapped to that position in the genome. A “genome of coverage of 30×” means that on average, each base of the genome has 30 reads mapped.
Sequence read	Inferred nucleotide sequence of a genome fragment. Reads range from short ( $\leq 300$ bp) to long (5 to 100+ kbp).
Contig	A contiguous sequence created by assembling multiple overlapping sequence reads.
Genome assembly	A singular (complete) or set of contigs after aligning and merging all sequence reads. Assemblies can be created <i>de novo</i> (relies only on sequence reads) or by mapping to a reference strain.
First-generation sequencing	Nucleotide sequencing that relied on either chain termination (Sanger) or cleavage (Maxam-Gilbert) methodology in single-tube reactions.
Second-generation sequencing	Nucleotide sequencing methods that sheared the genome and PCR amplified individual DNA fragments to massively parallelize sequencing and detect base identity by monitoring release of pyrophosphate (454), release of hydrogen (Ion Torrent), release of fluorescent reversible-terminator nucleotides (Illumina), or fluorescent ligated probes (SOLiD).
Third-generation sequencing	Nucleotide sequencing that relies on real-time single-molecule sequencing via monitoring of fluorescently labeled nucleotide incorporation (Pacific Biosciences) or ion current after DNA is fed through a channel (Oxford Nanopore).

genomes from animals and humans was searched to find four previously undiagnosed colistin-resistant isolates.

### NEXT-GENERATION SEQUENCING TECHNOLOGIES DRIVING WGS-AST

Two waves of sequencing technology innovation, the “second generation” starting in the mid-2000s and the “third generation” from about 2010 on, transformed all aspects of genomics and set the stage for WGS-AST (13, 14) (Table 1 and Fig. 1). Second-generation instruments, represented by the currently dominant Illumina sequencing-by-synthesis technology, sharply reduced the cost to generate data (i.e., \$/Mb), which led to large-scale sequencing of thousands of pathogen genomes and the use of shotgun metagenomics for clinical diagnostics. Illumina sequence reads are short ( $\leq 300$  bp) (15), typically paired-end, and have low per-base error rates (typically  $< 0.1\%$ ). Illumina sequencing allows deep shotgun coverage with high consensus accuracy, but *de novo* assembly typically results in genomes fragmented into multiple contigs and collapsed repeat regions (Fig. 1).

Third-generation single-molecule sequencing, exemplified by Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) technologies, produces much longer reads (typically 5 to 100 kbp, but there are recent reports of reads of  $> 2$  Mb [16]). Genome assemblies generated with long reads have fewer gaps and often span lengthy repeat regions, allowing the resolution of complex structural features, such as tandem repeats and nested insertions (17). Third-generation technologies have higher cost per base and higher per-base error rates than Illumina (5 to 15%), although improvements to chemistry and base calling algorithms are reducing the differential (18, 19). Error rates can in part be compensated by increasingly achievable higher read depths. For example, recent ONT studies generated  $> 100\times$  genome coverage and consensus error rates of  $< 0.08\%$  (20, 21). This level of error is adequate for most WGS-AST methods based on gene detection but may still be too high for many single-nucleotide polymorphism (SNP)-based methods. Hybrid assemblies (22) combine the accuracy of Illumina data and the gap-free assembly of ONT/PacBio and can achieve accuracies of  $> 99.9\%$ , but they rely on creating multiple libraries per sample, hence increasing cost and data management complexity.

The ONT technology has particular potential for WGS-AST because DNA sequence data become available within minutes of starting the sequencing run. A number of clinical gene-based ASTs for ONT have been piloted (see, e.g., references 21 and 23–25). At least two publications have demonstrated “streaming” AST, where the antibiotic resistance profile is updated in real time as the sequencing data are produced by the

**TABLE 2** Types of antibiotic resistance loci

Locus type	Description <sup>a</sup>
Gene	Presence of an intact protein-coding gene that confers resistance. For example, a strain that contains <i>blaZ</i> is inferred to be resistant to beta-lactams.
Plasmid/mobile element	Presence of a known drug resistance plasmid or mobile genetic cassette (e.g., <i>SCCmec</i> [114]) is used to infer that <i>S. aureus</i> is resistant to a beta-lactams.
Mutation	A particular SNP or SNV (which encompasses both SNPs and 1-bp indels) that is associated with resistance.
Allele	Nucleotide variant of gene caused by mutation. One sequence variant of a gene may be sensitive to a drug, while another allele may be associated with resistance.
Gene amplification	Increase in gene copy number due to homologous recombination. For example, a single gene in a genome may be sensitive to a drug, but a strain with two or more tandem repeats may be resistant.

<sup>a</sup>*SCCmec*, staphylococcal cassette chromosome *mec* element; SNV, single-nucleotide variant.

instrument (20, 26). Brinda et al. (26) used an indirect lineage-based method for predicting the resistance phenotype, which has the potential to mitigate some of the disadvantages of low-coverage ONT data, especially low SNP calling accuracy.

The outlook for clinical sequencing constantly shifts as existing technologies mature and become more cost-effective and new approaches emerge. Currently, Illumina is the dominant platform for WGS-AST. However, the approximate minimum cost of \$80 per genome is still too high for routine use in clinical laboratories. Minimum costs are based on 1-week turnaround times and batching many samples together for efficiency. Obtaining results within <24 h for small numbers of samples is possible but comes at a price penalty. Further decreases in cost and turnaround times will be needed before sequencing will be economically viable for routine WGS-AST. New instrumentation, such as the ONT Flongle disposable flow cells, might allow quicker and less expensive sequencing of smaller numbers of strains. Emerging synthetic long-read technologies (e.g., 10× Genomics) that rely on the barcoding of long DNA fragments to associate the resulting short reads during assembly (27) may provide the combination of read length, accuracy, and reduced cost that opens the door for routine WGS-AST and eventually routine metagenomic-AST.

### WGS-AST BASED ON SEARCHING CATALOGS OF RESISTANCE LOCI

At its essence, WGS-AST attempts to estimate the phenotype that would have been ascertained if the strain were subjected to a gold standard culture-based antibiotic resistance test. The simplest approaches seek to classify the strain as either susceptible or resistant to specific antibiotics, as defined by CLSI or EUCAST guidelines (3). More complex models have further tried to predict the MIC of an antibiotic for the strain. The most straightforward approach is to use a “rules-based” classification based on the presence of one or more known antimicrobial resistance (AMR) genes or mutations (Tables 2 and 3). This requires cross-referencing the genome sequence against databases of antibiotic resistance determinants. Databases have been developed mostly from curation of the literature on molecular genetic studies that link antibiotic resistance phenotypes to genes (28). Multispecies databases include CARD (29, 30), ResFinder (31) and its companion PointFinder (32), ARG-ANNOT (33), ARDB (34), MEGARes (35), Resfams (36), RAST (37), and the Bacterial Antimicrobial Resistance Reference Gene Database (BARRGD; <https://www.ncbi.nlm.nih.gov/bioproject/313047>). In addition, there are databases developed for single organisms, such as Dream TB (38) and MUBII-TB-DB (39) for *M. tuberculosis*.

Software tools for rules-based matching of antibiotic resistance catalogs operate on data produced at two points in the workflow for next-generation sequencing, raw sequence data and assembled contigs (Fig. 1). Each has tradeoffs in terms of speed of result and accuracy.

Detecting resistance in raw reads obviates the need for assembly and can therefore reduce the time to result if the algorithms are efficient. However, false positives may be introduced because of sequencing errors present in individual reads or DNA contamination from other organisms. Setting minimum thresholds for the number of

**TABLE 3** Selected rule-based WGS-AST results<sup>a</sup>

Species	Antibiotic(s)	No. of genomes tested	Diversity/no. of STs	Primary database(s)	Software	Input data	Sensitivity (%)	Specificity (%)	Reference
<i>E. coli</i>	Amoxicillin-clavulanate	76	NR	Custom	Blastx, ClustalW	Assembly, FASTQ	100	100	Tyson et al. (71)
	Trimethoprim	48	19+ STs	ResFinder	ResFinder	Assembly or FASTQ	100	100	Zankari et al. (115)
	Gentamicin	74	NR	>100 loci	Blastn	Assembly	100	100	Stoesser et al. (70)
	Pyrazinamide	167	NR	PhyResSE	Stampy	FASTQ	88.9	100, 94.9 (with uncharacterized)	Pankhurst et al. (64)
<i>M. tuberculosis</i>	Isoniazid	693	Lineages 1–4	TBDReaMDB, MUBil-TB-DB minus phylogenetic SNPs	TB Profiler	FASTQ	92.8	100	Coll et al. (63)
	Moxifloxacin	13,424	NR	NR	NR	NR	88.2	90	Miotto et al. (74)
	Amikacin	667	7 clades	Custom	SAMtools, mpileup, Cortex	FASTQ	91.2, 88.1 (with uncharacterized)	99.4, 99.5 (with uncharacterized)	Walker et al. (116)
	Rifampin	1,565	NR	Hain, Cepheid, AID <sup>b</sup> assays, literature	Mykrobe	FASTQ	90.8	99	Bradley et al. (5)
<i>S. aureus</i>	Ethambutol	752	NR	Mykrobe	Mykrobe	FASTQ	100	98.5, 77.3 (with uncharacterized)	Quan et al. (72)
	Fusidic acid	491	61 STs	Custom	Blastn, tblastn	Assembly	91	100	Gordon et al. (66)
	Vancomycin	NR	16 CCs	Custom	Blastn, mapping software	Assembly, FASTQ	100	100	Aanensen et al. (67)
	Mupirocin	340	25 CCs	Modified from Gordon et al. (66)	Mykrobe	FASTQ	100	100	Bradley et al. (5)
Non-serovar Typhi <i>S. enterica</i>	Ciprofloxacin, clindamycin, erythromycin, fusidic acid, gentamicin, methicillin, mupirocin, penicillin, rifampin, tetracycline, trimethoprim, vancomycin	1,379	111 STs	Custom	Mykrobe, GeneFinder, Typewriter	FASTQ, Assembly	97	99	Mason et al. (42)
	Chloramphenicol	332	ST1 and ST2	CARD, ResFinder, literature	GeneFinder	FASTQ	100	100	Day et al. (61)
	Ceftriaxone	640	NR	Tyson et al. (71)	Blastx, ClustalW	Assembly, FASTQ	100	99.8	McDermott et al. (59)
	Ciprofloxacin	3,491	227 serovars	CARD, ResFinder	GeneFinder	FASTQ	99.28	99.97	Neuert et al. (60)
<i>S. pneumoniae</i>	Erythromycin	210	90 STs	SRST2	SRST2	FASTQ	100	100	Deng et al. (69)
	<i>Campylobacter jejuni</i> , <i>Campylobacter coli</i>	32/82	NR	Tyson et al. (71), ARDB, ResFinder	Blastx, ClustalW	Assembly, FASTQ	100	100	Zhao et al. (62)
<i>Enterococcus faecalis</i> , <i>Enterococcus faecium</i>	Kanamycin	50	12 STs	ResFinder	ResFinder	Assembly or FASTQ	100	100	Zankari et al. (115)
	Levofloxacin	390	175 STs	Custom	NR	Assembly	91.9	93.7	Kos et al. (68)
<i>Pseudomonas aeruginosa</i> , <i>Klebsiella pneumoniae</i> , <i>Shigella sonnei</i>	Gentamicin	69	NR	Custom	Blastn	Assembly	96	100	Stoesser et al. (70)
	Ampicillin	341	NR	CARD, ResFinder	GeneFinder	FASTQ	100	100	Sadokki et al. (65)

<sup>a</sup>NR, not reported; ST, sequence type; CC, clonal complex.

<sup>b</sup>AID, Autoimmun Diagnostika GmbH.

reads needed for a positive result can help overcome read error problems. Software tools use different strategies for processing raw reads. KmerResistance (40) matches k-mer subsequences of the resistance locus catalog against raw reads similarly split into the same length subsequence. SRST2 (41) and GeneFinder (42) use the efficient read alignment program Bowtie (43) to map genes to the read set as a first step before enumerating SNPs and gene matches. ARIBA (44) uses a partial *de novo* assembly after first recruiting individual reads that may map to target genes. The web tool Point-Finder (32) identifies known point mutations after mapping reads against reference genomes. Mykrobe (5) creates de Bruijn graphs of contigs from raw data and matches them against known genes; however, for speed, it omits generation of a consensus sequence (42).

Genome assembly can either be *de novo* or by mapping to a reference strain. *De novo* assembly usually produces more fragmented genome assemblies but avoids the biases of building the assembly on an existing reference template. For reference-based assembly, single nucleotide polymorphism (SNP) detection becomes less accurate the greater the distance between isolate and reference, but *de novo*-assembled DNA sequences are free from this bias. However, an antibiotic resistance gene may be missed when using a *de novo* assembly if it is split across multiple contigs. Multicopy genes associated with antibiotic resistance, such as rRNA, present a particular challenge. A mutation in just one or two copies is sometimes sufficient to impart a resistance phenotype (e.g., in the case of azithromycin resistance caused by mutations in *Neisseria* 23S rRNA [45]), yet assembly algorithms seek a consensus sequence; thus, important genetic variants within repeats could be missed if repeats “collapse” into a single copy (3, 46). Using longer reads as input for assembly can overcome the problem of collapsed repeats. Increasing sequence coverage should also minimize these types of errors, but more data come with increased cost and slower computation time. This can be partially addressed by downsampling coverage, as after about 100× Illumina genome coverage of *Staphylococcus aureus* genomes, adding more data to the assembly produced little extra benefit for assembly accuracy at the cost of significant declines in processing speed (47). Most catalog-based software programs that take assembled data as input (e.g., Typewriter [42], SSTAR [48], CARD RGI [30], ARG-ANNOT [33], ResFinder [31], and ABRicate [<https://github.com/tseemanNABricate>]) use some form of BLAST alignment and results parsing, which usually takes a small fraction of the processor time used to construct the *de novo* assembly.

For many species and antibiotic resistance phenotypes, there is good concordance between what is known about the genetic basis of resistance and the resistance phenotype. Rules-based WGS-AST has been shown to have high sensitivity and specificity (>95%) for many phenotypes across several pathogen species (Table 3) (5, 32, 40, 42, 49–74), although with the caveat that studies varied widely in the number of strains tested and the within-species genetic diversity of the test set. However, there were some cases, e.g., levofloxacin resistance in *P. aeruginosa* (Table 3) (68), where sensitivity and specificity were below 95%. In an extensive survey of the genetic basis of resistance in *M. tuberculosis*, Miotto et al. (74) classified the predictive power of *M. tuberculosis* mutations into high, moderate, and minimal confidence. Therefore, rules-based approaches alone may not always be sufficient for accurate WGS-AST.

### MODEL-BASED ANTIBIOTIC RESISTANCE PREDICTION

Most rules-based methods make a number of (often unacknowledged) assumptions about the phenotypes they attempt to predict. These assumptions include (i) that either a single genetic locus is responsible for the phenotype, or, if multiple loci are present, that they do not interact in a complex manner (i.e., absence of epistasis [75–77]); (ii) that loci are highly penetrant and are not affected by the strain background; and (iii) that there is complete knowledge of the genetic basis of the phenotype. For a large number of cases, these assumptions do not completely hold. Some studies have attempted to capture uncertainty in the genetic basis of resistance and reduce overfitting using a variety of statistical modeling and machine learning (ML)

approaches (Table 4) (45, 78–85). For simplicity, we have placed them together here under the term “model-based” prediction.

The most common strategy used is to train a classifier on a set of genomes with known phenotypes. The classifier can be asked to learn which SNPs, indels, or other genetic features are important for the phenotype *ab initio* or be given a set of features already known to be important based on existing databases or a combination of both. As not all AMR determinants contribute equally to the antibiotic resistance of a strain, noise in phenotype prediction can often be reduced and accuracy increased by weighting each locus using a machine learning model. Models can also be trained to take into account potential interactions between loci. The accuracy of the model is determined by predicting resistance in a second set of phenotyped genomes, ideally those from strains different from the training set. Models can be used to predict either sensitivity or resistance based on an accepted threshold or the MIC of the strain to the particular antibiotic. As for rules-based methods, data inputs can be reads, k-mers, and assembled contigs (Table 4). Time to result is highly variable and is dependent on factors, such as the number of features used and the complexity of the ML algorithm. Three examples, discussed briefly below, illustrate different model-based approaches to prediction and some of the advantages and pitfalls.

Yang et al. (79) examined the genomes of 1,839 *M. tuberculosis* strains resistant to eight drugs isolated in the UK, using reference mapping to identify mutations in 23 putative resistance genes identified in earlier experimental studies. Because of the lack of lateral gene transfer in *M. tuberculosis*, resistance arises primarily through mutations (in this case, SNPs). Seven ML models were built and compared to simple rules-based models, where the presence of a known mutation indicated resistance. Different subsets of mutations were also tested to determine the effect on performance. The best ML method for each drug increased sensitivity over the rules-based models by 2 to 24%, with the trade-off of minor losses in specificity due to strains labeled as “susceptible” containing mutations associated with resistance. However, it is possible that the phenotypes of these strains were determined incorrectly. The best model for each drug varied in both the ML algorithm used and the mutation subset used, indicating that there is no one-size-fits-all solution, and the model chosen for prediction can be optimized for the complexity of resistance phenotype and amount of *a priori* knowledge.

In an alternative approach to the multiple-locus problem, Eyre et al. (45) predicted *Neisseria gonorrhoeae* MICs to cefixime, penicillin, azithromycin, ciprofloxacin, and tetracycline using multivariate linear regression. Multiple genetic loci (SNPs, plasmids, and alleles of the *penA* gene) were already known to make contributions to resistance. The presence or absence of each candidate locus was ascertained in 681 *N. gonorrhoeae* genomes by mapping reads to a reference genome. Backwards selection (where each variable is removed until the information lost, as estimated by the Akaike information criterion score, is minimized) was used to reduce the number of loci in the model and limit overfitting. Overall, model-fitted MICs were within two doubling dilutions of the MIC for 98% of the strains. Using EUCAST cutoffs, the sensitivity of calling resistance was 98.7%, and the specificity was 98.3%.

The final example of successful implementation of model-based methods is the prediction of beta-lactam resistance in *Streptococcus pneumoniae* (80, 86). Most of the strain-specific variation in levels of resistance in *S. pneumoniae* was found to be driven by amino acid sequence variation in the transpeptidase domain of three penicillin binding proteins (PBPs). Proteins with similar sequence signatures in their transpeptidase domains were clustered into “PBP types.” There were high levels of horizontal transfer of genes in the *S. pneumoniae* species, but the strain genomic background outside the PBPs only contributed 1 to 6% of the variation in MIC (86); instead, the PBP type was the most important variable. In 4,309 *Streptococcus* genome sequences where the PBP type was ascertained from *de novo*-assembled data (80, 87), the MIC could be accurately predicted using both a rules-based model (“Mode MIC,” where the MIC was the most frequent in the closest known PBP type) and a machine learning (Random

**TABLE 4** Selected model-based WGS-AST results

Species	Antibiotic(s)	No. of genomes tested	Diversity <sup>a</sup>	Database <sup>b</sup>	ML algorithm	Input data	Sensitivity (%)	Specificity (%)	Overall accuracy (%) <sup>c</sup>	Reference
<i>E. coli</i>	Amoxicillin	329	7 STs	NA	Gradient-boosted trees	Pangenome, population structure matrix	90	95		Moradigarvand et al. (78)
	Ciprofloxacin	581	7 STs	NA	Gradient-boosted trees	Pangenome, population structure matrix, SNPs	81	99		Moradigarvand et al. (78)
<i>M. tuberculosis</i>	Gentamicin	564	7 STs	NA	Gradient-boosted trees	Pangenome, population structure matrix	87	99		Moradigarvand et al. (78)
	Trimethoprim	283	7 STs	NA	Gradient-boosted trees	Pangenome, population structure matrix	92	97		Moradigarvand et al. (78)
	Isoniazid	1,811 (80% train, 20% test)	7 clades	NA	Random Forest	Variants in 23 genes	97	94		Yang et al. (79)
	Rifampin	1,725 (80% train, 20% test)	7 clades	NA	Class-conditional Bernoulli mixture model	Variants in 23 genes	97	97		Yang et al. (79)
<i>S. pneumoniae</i>	Ethambutol	3,526 (80% train, 20% test)	5 genetic clusters	NA	Multitask wide and deep neural networks	Variants in 32 regions	91.9	90.3		Chen et al. (81)
	Pyrazinamide	3,147 (train), 567 (test)	5 genetic clusters	NA	Multitask wide and deep neural networks	Variants in 32 regions	75.2	90.1		Chen et al. (81)
	Kanamycin	162 (train), 18 (test)	NR	PATRIC, RAST	AdaBoost	Assembly		88.3 (F1)		Davis et al. (83)
<i>S. pneumoniae</i>	Beta-lactams (PEN, AMO, MER, TAX, CFT, CFXY) <sup>d</sup>	2,528 (train), 1,781 (test)	403 STs (train), 299 STs (test)	NA	Random Forest	PBP sequences				Li et al. (80)
	Beta-lactams	1,350 (train), 58 (test)	NR	PATRIC, RAST	AdaBoost	Assembly				Davis et al. (83)
<i>N. gonorrhoeae</i>	Azithromycin	681	NR	NA	Linear regression	Variants in 20 regions	80, 99 (±1 MIC dilution)	83, 94 (±1 MIC dilution)	93 (±1 MIC dilution), 44 (category)	Eyre et al. (45)
	Ciprofloxacin	676	NR	NA	Linear regression	Variants in 20 regions	100	99	94 (±1 MIC dilution), 68 (category)	Eyre et al. (45)
<i>K. pneumoniae</i>	Ampicillin-sulbactam	1,668	>99 STs	PATRIC, RAST	XGBoost	Assembly				Nguyen et al. (117)
	Levofloxacin	1,668	>99 STs	PATRIC, RAST	XGBoost	Assembly				Nguyen et al. (117)
	Meropenem	1,777	>99 STs	PATRIC, RAST	AdaBoost	Assembly				Long et al. (118)
	Piperacillin-tazobactam	1,777	>99 STs	PATRIC, RAST	AdaBoost	Assembly				Long et al. (118)
<i>E. coli</i> , <i>Enterobacter aerogenes</i> , <i>K. pneumoniae</i>	Ampicillin	78	NR	Resfams	Logistic regression	FASTA, alignments			97.4	Peseky et al. (85)
	Chloramphenicol	78	NR	Resfams	Logistic regression	FASTA, alignments			89.7	Peseky et al. (85)
<i>S. aureus</i>	Methicillin	99 (train), 11 (test)	NR	PATRIC, RAST	AdaBoost	Assembly			99.5 (F1)	Davis et al. (83)
	Vancomycin	75	12 STs	Custom	Random Forest	Assembly	73		81	Alam et al. (82)
<i>Acinetobacter baumannii</i> , Non-serovar Typhi <i>S. enterica</i>	Carbapenem	99 (train), 11 (test)	NR	PATRIC, RAST	AdaBoost	Assembly			95 (F1)	Davis et al. (83)
	Ceftriaxone	5,278	NR	PATRIC, RAST	XGBoost	Assembly			80, 95 (±1 MIC dilution)	Nguyen et al. (84)

<sup>a</sup>ST, sequence type; NR, not reported.  
<sup>b</sup>NA, not applicable.  
<sup>c</sup>F1, harmonic average of the precision (positive predictive value [PPV]) and recall (sensitivity) (84).  
<sup>d</sup>PEN, penicillin; AMO, amoxicillin; MER, meropenem; TAX, ceftriaxone; CFT, cefotaxime; CFY, cefixime.



Forest) classifier using the amino acid sequence of the transpeptidase domains. Mode MIC and Random Forest false-positive rates were <3%, but critically, while the false-negative rates for the Random Forest classifier were low, the rates for Mode MIC were above 50% for some beta-lactams. The reason for the poor performance of the rule-based prediction was the misclassification of a subset of PBP types with significant sequence divergence from the training set. The good performance of the Random Forest classifier across previously unseen sequences was because the machine learning method discovered the previously unappreciated significance for resistance of a group of key amino acid residues in the transpeptidase domains supplied in the training set.

### CONSEQUENCES OF INCORRECT AST PREDICTION AND METHODS TO INCREASE ROBUSTNESS

Errors in the sensitivity and specificity of genome prediction of antimicrobial phenotypes, either false positives (phenotypically susceptible, WGS-prediction resistant) or false negatives (phenotypically resistant, WGS-prediction susceptible) have different consequences for treatment. False negatives are considered most concerning, as they can lead to inadequate treatment of a resistant infection, increasing morbidity and mortality. It is often preferable to reduce false negatives at the cost of increasing the false-positive rate (66), although false-positive results lead to inappropriate antibiotic use, potentially harming the patient and increasing the risk of resistance to last-line antibiotics.

How can the accuracy of genetic prediction be improved in the future? It is important in development of WGS-AST tools to have as large and diverse of a test and training strain set as possible. Most studies to date have used a convenience sample of strains based on accessible collections with limited geographic and temporal variability. Even if large numbers of strain genomes are obtained, accuracy statistics can be misleading if isolates from locally abundant clonal lineages are heavily overrepresented. In reporting the isolate collection used for WGS-AST development and testing, we need to develop accepted statistics for the assessment of genetic diversity. These could be pairwise average nucleotide identity, number of multilocus sequence type (MLST)/cgMLST, a sequence entropy measure, or the percentage of known species pangenome represented. Epistatic effects could explain the reduced accuracy of WGS-AST across diverse strain backgrounds, especially those not included in test sets. The relatively few studies on epistasis and antibiotic resistance suggest that few generalizations can be made across species and phenotypes. Knopp and Andersson (77) found little variation in phenotypic expression for 13 resistance mutations across 10 strains of *Salmonella enterica* and *Escherichia coli* (even when as many as four mutations were combined in one background). However, there are reports of epistatic effects both between resistance mutations and between the resistance mutation and the strain genetic background in *Pseudomonas aeruginosa* and other species (75, 88, 89). Extensive empirical testing will be needed to understand how resistance phenotype expression varies across diverse strain backgrounds in each species. New statistical methods coupled with very large pathogen genomic data sets can also lead to the discovery of novel epistatic interactions (90).

The finding that rules-based methods for predicting beta-lactam MICs from PBP type performed poorly on previously unseen strain types (80), discussed above, is an example of how incomplete knowledge impacts WGS-AST. We do not yet understand the genetic basis of resistance for some antibiotics. For others, even where there are well-defined canonical mutations or genes, it is becoming apparent that there are a large number of rare genetic events that could lead to resistance. Guérillot et al. (91) found through deep sequencing of lab-selected colony pools that while eight *rpoB* mutations were responsible for 93% of the rifampin resistance in *S. aureus* clinical strains (92), the remaining 8% were caused by least 72 rarer mutations (91). Wistrand-Yuen et al. (93) showed that selection for streptomycin resistance in *Salmonella enterica* at sub-MIC levels resulted in strains with an array of novel low-effect mutations different from classic high-effect drug resistance. Potentially even harder to predict

based on sequence alone are resistance phenotypes caused by rare genetic events, such as insertion sequence movement into or upstream of open reading frames (94, 95).

In the face of this genetic complexity, how can WGS-AST continue to improve? It is unrealistic to expect 100% accuracy for every drug. We may be able to establish a stable level of uncertainty for each phenotype and learn to accept that some are more genetically plastic than others. Improvements may also be possible by incorporating functional prediction algorithms for mutations in resistance genes where the functional effect is untested. At its simplest, this could be taking into account frameshift mutations that, for example, in a *blaZ* gene would be predicted to inactivate translation of the beta-lactamase and yield a phenotypically susceptible strain. More sophisticated metrics include delta-bitscore, which uses hidden Markov models to identify functional divergence in orthologous genes and is an improvement on the previous use of the ratio of nonsynonymous to synonymous evolutionary changes ( $dN/dS$ ) (used to indicate positive selection), which was prone to false positives (96). *S. aureus* resistance to trimethoprim has been accurately predicted based on the free energy state of the dihydrofolate reductase amino acid sequence (97), suggesting an avenue for sophisticated structural prediction studies. One straightforward way to increase robustness of WGS-AST prediction is to combine predictions of multiple tools that examine different facets of the WGS data set (i.e., k-mers and assembled contigs) (42).

### THE PHENOTYPE PROBLEM

As mentioned earlier, WGS-AST uses culture-based antibiotic phenotypes as the gold standard, but culture-based methods are not themselves free of uncertainty. Culture-based assays for drugs in some species have been shown to be unreliable, for example, ethambutol, pyrazinamide, and rifampin in *M. tuberculosis* (98–101). Even for drug resistances that can be tested reliably, there is evidence to suggest that the error in culture-based AST between different reference laboratories is greater than the inaccuracy of WGS-AST (45, 57, 67). Errors may arise from subtle operator-specific biases or variations between individual automated instruments that perform rapid organismal identification and susceptibility testing. Additionally, unlike standard research laboratory experimentation, it is generally not practical to perform replicate MIC tests in clinical laboratories.

Culture-based AST measures resistance in a highly constrained setting that does not reflect the variance of expression of the phenotype, especially during human infections. Many culture-based MICs are not robust to changes in growth medium, bacterial cell density, or temperature. Antimicrobial susceptibility may be altered due to the formation of biofilms and/or the synergistic or antagonistic interactions with other bacterial species (102–104). Resistance gene expression can be induced to higher levels *in vivo* than *in vitro* (105), small resistant subpopulations (heteroresistant strains) or mixed infections can quickly exhibit phenotypic resistance when exposed to selection, and persisters (106), which are not detected by AST, can survive treatment and cause recurrent infections. This underlies the fundamental issue of using MIC breakpoints to define resistance. Breakpoints are not based on genetic correlates of resistance but on factors, such as distribution of MICs, pharmacokinetics of the drug, and clinical treatment experience, among others. Consequently, there is considerable debate on how to set breakpoints, which can vary across countries and organizations (107). Breakpoints are often revised due to new epidemiological data, but adoption can sometimes be slow, and changes can serve to confound retrospective meta-analyses of antibiotic resistance trends and mechanisms defined in the literature.

Despite culture-based AST being an imperfect gold standard for WGS-AST, it is unlikely that it will soon be replaced. However, due to a revolution in NAAT-based typing, clinical laboratories are culturing many fewer isolates, especially hard-to-grow species, such as *Legionella* spp. and *M. tuberculosis*, or sexually transmitted diseases, such as those caused by *Neisseria* spp., which need a fast turnaround time for diagnosis. As sequencing becomes less expensive and more accurate, metagenomic-AST (Fig. 1)

**TABLE 5** Some outstanding questions

Question
At what price and turnaround time will WGS-AST replace culture-based sequencing for routine use in clinical microbiology labs?
How do we interpret the presence of an antimicrobial resistance determinant gene if the susceptibility of the strain is below the MIC?
Can genome prediction be used to detect heteroresistance? Or to detect polygenic phenotypes?
How important is epistasis in determining the resistance to different classes of antibiotics?
Can gene amplification as a mechanism of resistance be accurately determined from WGS data?
How efficiently can WGS-AST prediction software be ported to metagenomic-AST data?

may render culturing of isolates unnecessary for routine treatment. However, as bacterial pathogens continue to evolve, it is possible that new resistance loci will emerge or the relationship between known AMR determinants and phenotypic resistance will change due to epistatic interactions with novel emerging strain backgrounds. Therefore, even when phenotypic testing is no longer used as a primary screen, it should continue to be used as quality control step to ensure that contemporary clinical strains have not drifted away from predicted phenotypes. Large multisite surveys that integrate rational strain sampling and centralized gold standard AST, such as the recent European *N. gonorrhoeae* study (57), will be needed to continually update and challenge existing WGS-AST methods.

### CONCLUDING REMARKS

The adoption of routine use of whole-genome sequencing in clinical microbiology would be a revolution in clinical medicine. While this review only discusses bacterial pathogens, there is also progress being made to use WGS to predict drug resistance in viruses, fungi, and eukaryotic parasites (108–110). The major barriers to adoption for routine use stem from technological limitations that make the process currently too slow and expensive and the need to demonstrate the advantages of a new approach to diagnosis for skeptical end-users in clinical microbiology laboratories (Table 5). Along with the development of instrumentation, there is need for efficient data management of WGS data, user-friendly pipelines that do not require specialized bioinformaticians for use, and rapidly interpretable results (111). As has been covered in more depth in other articles (3, 4, 112), standardization of software, sequence data, genomic test sets and phenotypes, and data storage are all needed to transform what is now a research enterprise into a robust clinical tool acceptable to both clinical microbiologists and regulatory agencies.

### ACKNOWLEDGMENTS

We thank Claudia Solis Lemus for reading and commenting on the draft manuscript.

T.D.R. was supported by National Institute of Allergy and Infectious Diseases (NIAID) award AI121860. M.S. was supported in part by the Antimicrobial Resistance and Therapeutic Discovery Training Program funded by NIAID T32 award AI106699-05.

The contents of this article are solely the responsibility of the authors and do not necessarily reflect the official views of the National Institutes of Health or the U.S. Government.

### REFERENCES

- Centers for Disease Control and Prevention. 2013. Antibiotic resistance threats in the United States. Centers for Disease Control and Prevention, Atlanta, GA. <https://www.cdc.gov/drugresistance/pdf/ar-threats-2013-508.pdf>.
- van Belkum A, Dunne WM, Jr. 2013. Next-generation antimicrobial susceptibility testing. *J Clin Microbiol* 51:2018–2024. <https://doi.org/10.1128/JCM.00313-13>.
- Ellington MJ, Ekelund O, Aarestrup FM, Canton R, Doumith M, Giske C, Grundman H, Hasman H, Holden MTG, Hopkins KL, Iredell J, Kahlmeter G, Köser CU, MacGowan A, Mevius D, Mulvey M, Naas T, Peto T, Rolain J-M, Samuelsen Ø, Woodford N. 2017. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST subcommittee. *Clin Microbiol Infect* 23:2–22. <https://doi.org/10.1016/j.cmi.2016.11.012>.
- Goldberg B, Sichtig H, Geyer C, Ledebner N, Weinstock GM. 2015. Making the leap from research laboratory to clinic: challenges and

- opportunities for next-generation sequencing in infectious disease diagnostics. *mBio* 6:e01888-15.
5. Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, Earle S, Pankhurst LJ, Anson L, de Cesare M, Piazza P, Votintseva AA, Golubchik T, Wilson DJ, Wyllie DH, Diel R, Niemann S, Feuerriegel S, Kohl TA, Ismail N, Omar SV, Smith EG, Buck D, McVean G, Walker AS, Peto TEA, Crook DW, Iqbal Z. 2015. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nat Commun* 6:10063. <https://doi.org/10.1038/ncomms10063>.
  6. Doyle RM, Burgess C, Williams R, Gorton R, Booth H, Brown J, Bryant JM, Chan J, Creer D, Holdstock J, Kunst H, Lozewicz S, Platt G, Romero EY, Speight G, Tiberi S, Abubakar I, Lipman M, McHugh TD, Breuer J. 2018. Direct whole-genome sequencing of sputum accurately identifies drug-resistant *Mycobacterium tuberculosis* faster than MGIT culture sequencing. *J Clin Microbiol* 56:e00666-18.
  7. Votintseva AA, Bradley P, Pankhurst L, Del Ojo Elias C, Loose M, Nilgiriwala K, Chatterjee A, Smith EG, Sanderson N, Walker TM, Morgan MR, Wyllie DH, Walker AS, Peto TEA, Crook DW, Iqbal Z. 2017. Same-day diagnostic and surveillance data for tuberculosis via whole genome sequencing of direct respiratory samples. *J Clin Microbiol* 55:1285–1298. <https://doi.org/10.1128/JCM.02483-16>.
  8. Feijao P, Yao H-T, Fornika D, Gardy J, Hsiao W, Chauve C, Chindelevitch L. 10 January 2018. MentaLiST—a fast MLST caller for large MLST schemes. *Microb Genom* <https://doi.org/10.1099/mgen.0.000146>.
  9. Chen L, Zheng D, Liu B, Yang J, Jin Q. 2016. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res* 44:D694–D697. <https://doi.org/10.1093/nar/gkv1239>.
  10. Gardy JL, Loman NJ. 2018. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet* 19:9–20. <https://doi.org/10.1038/nrg.2017.88>.
  11. Falgenhauer L, Waezsada S-E, Yao Y, Imirzalioglu C, Käsbohrer A, Roesler U, Michael GB, Schwarz S, Werner G, Kreienbrock L, Chakraborty T, RESET Consortium. 2016. Colistin resistance gene *mcr-1* in extended-spectrum  $\beta$ -lactamase-producing and carbapenemase-producing Gram-negative bacteria in Germany. *Lancet Infect Dis* 16:282–283. [https://doi.org/10.1016/S1473-3099\(16\)00009-8](https://doi.org/10.1016/S1473-3099(16)00009-8).
  12. Schürch AC, van Schaik W. 2017. Challenges and opportunities for whole-genome sequencing-based surveillance of antibiotic resistance. *Ann N Y Acad Sci* 1388:108–120. <https://doi.org/10.1111/nyas.13310>.
  13. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. 2017. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* 35:833–844. <https://doi.org/10.1038/nbt.3935>.
  14. Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 17:333–351. <https://doi.org/10.1038/nrg.2016.49>.
  15. Reinert K, Langmead B, Weese D, Evers DJ. 2015. Alignment of next-generation sequencing reads. *Annu Rev Genomics Hum Genet* 16:133–151. <https://doi.org/10.1146/annurev-genom-090413-025358>.
  16. Payne A, Holmes N, Rakyen V, Loose M. 2018. Whale watching with BulkVis: a graphical viewer for Oxford Nanopore bulk fast5 files. *bioRxiv* <https://doi.org/10.1101/312256>.
  17. Giordano F, Aigrain L, Quail MA, Coupland P, Bonfield JK, Davies RM, Tischler G, Jackson DK, Keane TM, Li J, Yue J-X, Liti G, Durbin R, Ning Z. 2017. De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms. *Sci Rep* 7:3935. <https://doi.org/10.1038/s41598-017-03996-z>.
  18. Rhoads A, Au KF. 2015. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* 13:278–289. <https://doi.org/10.1016/j.gpb.2015.08.002>.
  19. Lu H, Giordano F, Ning Z. 2016. Oxford Nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics* 14:265–279. <https://doi.org/10.1016/j.gpb.2016.05.004>.
  20. Cao MD, Ganesamoorthy D, Elliott AG, Zhang H, Cooper MA, Coin LJM. 2016. Streaming algorithms for identification of pathogens and antibiotic resistance potential from real-time MinION sequencing. *Gigascience* 5:32. <https://doi.org/10.1186/s13742-016-0137-2>.
  21. Lemon JK, Khil PP, Frank KM, Dekker JP. 2017. Rapid Nanopore sequencing of plasmids and resistance gene detection in clinical isolates. *J Clin Microbiol* 55:3530–3543. <https://doi.org/10.1128/JCM.01069-17>.
  22. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
  23. Schmidt K, Mwaigwisya S, Crossman LC, Doumith M, Munroe D, Pires C, Khan AM, Woodford N, Saunders NJ, Wain J, O'Grady J, Livermore DM. 2016. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J Antimicrob Chemother* 72:104–114.
  24. Judge K, Hunt M, Reuter S, Tracey A, Quail MA, Parkhill J, Peacock SJ. 2016. Comparison of bacterial genome assembly software for MinION data and their applicability to medical microbiology. *Microb Genom* 2:e000085.
  25. van der Helm E, Imamovic L, Hashim Ellabaan MM, van Schaik W, Koza A, Sommer MOA. 2017. Rapid resistome mapping using nanopore sequencing. *Nucleic Acids Res* 45:e61.
  26. Břinda K, Callendrello A, Cowley L, Charalampous T, Lee RS, MacFadden DR, Kucherov G, O'Grady J, Baym M, Hanage WP. 2018. Lineage calling can identify antibiotic resistant clones within minutes. *bioRxiv* <https://doi.org/10.1101/403204>.
  27. Kuleshov V, Snyder MP, Batzoglou S. 2016. Genome assembly from synthetic long read clouds. *Bioinformatics* 32:i216–i224. <https://doi.org/10.1093/bioinformatics/btw267>.
  28. Xavier BB, Das AJ, Cochrane G, De Ganck S, Kumar-Singh S, Aarestrup FM, Goossens H, Malhotra-Kumar S. 2016. Consolidating and exploring antibiotic resistance gene data resources. *J Clin Microbiol* 54:851–859. <https://doi.org/10.1128/JCM.02717-15>.
  29. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski AC, Piddock LJV, Spanogiannopoulos P, Sutherland AD, Tang I, Taylor PL, Thaker M, Wang W, Yan M, Yu T, Wright GD. 2013. The Comprehensive Antibiotic Resistance Database. *Antimicrob Agents Chemother* 57:3348–3357. <https://doi.org/10.1128/AAC.00419-13>.
  30. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, Doshi S, Courtot M, Lo R, Williams LE, Frye JG, Elsayegh T, Sardar D, Westman EL, Pawlowski AC, Johnson TA, Brinkman FSL, Wright GD, McArthur AG. 2017. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res* 45:D566–D573. <https://doi.org/10.1093/nar/gkw1004>.
  31. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 67:2640–2644. <https://doi.org/10.1093/jac/dks261>.
  32. Zankari E, Allesøe R, Joensen KG, Cavaco LM, Lund O, Aarestrup FM. 2017. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *J Antimicrob Chemother* 72:2764–2768. <https://doi.org/10.1093/jac/dkx217>.
  33. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, Rolain J-M. 2014. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother* 58:212–220. <https://doi.org/10.1128/AAC.01310-13>.
  34. Liu B, Pop M. 2009. ARDB—Antibiotic Resistance Genes Database. *Nucleic Acids Res* 37:D443–D447. <https://doi.org/10.1093/nar/gkn656>.
  35. Lakin SM, Dean C, Noyes NR, Dettenwanger A, Ross AS, Doster E, Rovira P, Abdo Z, Jones KL, Ruiz J, Belk KE, Morley PS, Boucher C. 2017. MEGARes: an antimicrobial resistance database for high throughput sequencing. *Nucleic Acids Res* 45:D574–D580. <https://doi.org/10.1093/nar/gkw1009>.
  36. Gibson MK, Forsberg KJ, Dantas G. 2015. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J* 9:207–216. <https://doi.org/10.1038/ismej.2014.106>.
  37. Antonopoulos DA, Assaf R, Aziz RK, Brettin T, Bun C, Conrad N, Davis JJ, Dietrich EM, Disz T, Gerdes S, Kenyon RW, Machi D, Mao C, Murphy-Oldson DE, Nordberg EK, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Santerre J, Shukla M, Stevens RL, VanOeffelen M, Vonstein V, Warren AS, Wattam AR, Xia F, Yoo H. 31 July 2017. PATRIC as a unique resource for studying antimicrobial resistance. *Brief Bioinform* <https://doi.org/10.1093/bib/bbx083>.
  38. Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB. 2009. Tuberculosis drug resistance mutation database. *PLoS Med* 6:e2. <https://doi.org/10.1371/journal.pmed.1000002>.
  39. Flandrois J-P, Lina G, Dumitrescu O. 2014. MUBII-TB-DB: a database of mutations associated with antibiotic resistance in *Mycobacterium tuberculosis*. *BMC Bioinformatics* 15:107. <https://doi.org/10.1186/1471-2105-15-107>.
  40. Clausen PTL, Zankari E, Aarestrup FM, Lund O. 2016. Benchmarking of methods for identification of antimicrobial resistance genes in bacterial

- whole genome data. *J Antimicrob Chemother* 71:2484–2488. <https://doi.org/10.1093/jac/dkw184>.
41. Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, Tomita T, Zobel J, Holt KE. 2014. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* 6:90. <https://doi.org/10.1186/s13073-014-0090-6>.
  42. Mason A, Foster D, Bradley P, Golubchik T, Doumith M, Gordon NC, Pichon B, Iqbal Z, Staves P, Crook D, Walker AS, Kearns A, Peto T. 2018. Accuracy of different bioinformatics methods in detecting antibiotic resistance and virulence factors from *Staphylococcus aureus* whole genome sequences. *J Clin Microbiol* 56:e01815-17.
  43. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
  44. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, Harris SR. 2017. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *bioRxiv* <https://doi.org/10.1101/118000>.
  45. Eyre DW, De Silva D, Cole K, Peters J, Cole MJ, Grad YH, Demczuk W, Martin I, Mulvey MR, Crook DW, Walker AS, Peto TEA, Paul J. 2017. WGS to predict antibiotic MICs for *Neisseria gonorrhoeae*. *J Antimicrob Chemother* 72:1937–1947. <https://doi.org/10.1093/jac/dkx067>.
  46. Sinclair A, Arnold C, Woodford N. 2003. Rapid detection and estimation by pyrosequencing of 23S rRNA genes with a single nucleotide polymorphism conferring linezolid resistance in Enterococci. *Antimicrob Agents Chemother* 47:3620–3622. <https://doi.org/10.1128/AAC.47.11.3620-3622.2003>.
  47. Petit RA, III, Read TD. 2018. *Staphylococcus aureus* viewed from the perspective of 40,000+ genomes. *PeerJ* 6:e5261. <https://doi.org/10.7717/peerj.5261>.
  48. de Man TJ, Limbago BM. 2016. SSTAR, a stand-alone easy-to-use antimicrobial resistance gene predictor. *mSphere* 1:e00050-15.
  49. Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, Ip CLC, Wilson DJ, Didelot X, O'Connor L, Lay R, Buck D, Kearns AM, Shaw A, Paul J, Wilcox MH, Donnelly PJ, Peto TEA, Walker AS, Crook DW. 2012. A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. *BMJ Open* 2 <https://doi.org/10.1136/bmjopen-2012-001124>.
  50. Köser CU, Holden MTG, Ellington MJ, Cartwright EJP, Brown NM, Ogilvy-Stuart AL, Hsu LY, Chewapreecha C, Croucher NJ, Harris SR, Sanders M, Enright MC, Dougan G, Bentley SD, Parkhill J, Fraser LJ, Betley JR, Schulz-Trieglaff OB, Smith GP, Peacock SJ. 2012. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N Engl J Med* 366:2267–2275. <https://doi.org/10.1056/NEJMoa1109910>.
  51. Leopold SR, Goering RV, Witten A, Harmsen D, Mellmann A. 2014. Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. *J Clin Microbiol* 52:2365–2370. <https://doi.org/10.1128/JCM.00262-14>.
  52. ElMaraachli W, Slater M, Berrada ZL, Lin S-YG, Catanzaro A, Desmond E, Rodrigues C, Victor TC, Crudu V, Gler MT, Rodwell TC. 2015. Predicting differential rifampicin resistance in clinical *Mycobacterium tuberculosis* isolates by specific *rpoB* mutations. *Int J Tuberc Lung Dis* 19:1222–1226. <https://doi.org/10.5588/ijtld.14.0936>.
  53. Lee GC, Long SW, Musser JM, Beres SB, Olsen RJ, Dallas SD, Nunez YO, Frei CR. 2015. Comparative whole genome sequencing of community-associated methicillin-resistant *Staphylococcus aureus* sequence type 8 from primary care clinics in a Texas community. *Pharmacotherapy* 35:220–228. <https://doi.org/10.1002/phar.1536>.
  54. Moran RA, Anantham S, Holt KE, Hall RM. 2017. Prediction of antibiotic resistance from antibiotic resistance genes detected in antibiotic-resistant commensal *Escherichia coli* using PCR or WGS. *J Antimicrob Chemother* 72:700–704. <https://doi.org/10.1093/jac/dkw511>.
  55. Ginn AN, Zong Z, Wiklendt AM, Thomas LC, Merlino J, Gottlieb T, van Hal S, Harkness J, Macleod C, Bell SM, Leroi MJ, Partridge SR, Iredell JR. 2013. Limited diversity in the gene pool allows prediction of third-generation cephalosporin and aminoglycoside resistance in *Escherichia coli* and *Klebsiella pneumoniae*. *Int J Antimicrob Agents* 42:19–26. <https://doi.org/10.1016/j.ijantimicag.2013.03.003>.
  56. Ginn AN, Wiklendt AM, Zong Z, Lin RTP, Teo JWP, Tambyah PA, Peterson LR, Kaul K, Partridge SR, Iredell JR. 2014. Prediction of major antibiotic resistance in *Escherichia coli* and *Klebsiella pneumoniae* in Singapore, USA and China using a limited set of gene targets. *Int J Antimicrob Agents* 43:563–565. <https://doi.org/10.1016/j.ijantimicag.2014.02.010>.
  57. Harris SR, Cole MJ, Spiteri G, Sánchez-Busó L, Golparian D, Jacobsson S, Goater R, Abudahab K, Yeats CA, Bercot B, Borrego MJ, Crowley B, Stefanelli P, Tripodo F, Abad R, Aanensen DM, Unemo M, Euro-GASP Study Group. 2018. Public health surveillance of multidrug-resistant clones of *Neisseria gonorrhoeae* in Europe: a genomic survey. *Lancet Infect Dis* 18:P758–P768.
  58. Holden MTG, Hsu L-Y, Kurt K, Weinert LA, Mather AE, Harris SR, Strommenger B, Layer F, Witte W, de Lencastre H, Skov R, Westh H, Zemlicková H, Coombs G, Kearns AM, Hill RLR, Edgeworth J, Gould I, Gant V, Cooke J, Edwards GF, McAdam PR, Templeton KE, McCann A, Zhou Z, Castillo-Ramírez S, Feil EJ, Hudson LO, Enright MC, Balloux F, Aanensen DM, Spratt BG, Fitzgerald JR, Parkhill J, Achtman M, Bentley SD, Nübel U. 2013. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res* 23:653–664. <https://doi.org/10.1101/gr.147710.112>.
  59. McDermott PF, Tyson GH, Kabera C, Chen Y, Li C, Folster JP, Ayers SL, Lam C, Tate HP, Zhao S. 2016. Whole-genome sequencing for detecting antimicrobial resistance in nontyphoidal *Salmonella*. *Antimicrob Agents Chemother* 60:5515–5520. <https://doi.org/10.1128/AAC.01030-16>.
  60. Neuert S, Nair S, Day MR, Doumith M, Ashton PM, Mellor KC, Jenkins C, Hopkins KL, Woodford N, de Pinna E, Godbole G, Dallman TJ. 2018. Prediction of phenotypic antimicrobial resistance profiles from whole genome sequences of non-typhoidal *Salmonella enterica*. *Front Microbiol* 9:592. <https://doi.org/10.3389/fmicb.2018.00592>.
  61. Day MR, Doumith M, Do Nascimento V, Nair S, Ashton PM, Jenkins C, Dallman TJ, Stevens FJ, Freedman J, Hopkins KL, Woodford N, De Pinna EM, Godbole G. 2017. Comparison of phenotypic and WGS-derived antimicrobial resistance profiles of *Salmonella enterica* serovars Typhi and Paratyphi. *J Antimicrob Chemother* 73:365–372.
  62. Zhao S, Tyson GH, Chen Y, Li C, Mukherjee S, Young S, Lam C, Folster JP, Whichard JM, McDermott PF. 2016. Whole-genome sequencing analysis accurately predicts antimicrobial resistance phenotypes in *Campylobacter* spp. *Appl Environ Microbiol* 82:459–466. <https://doi.org/10.1128/AEM.02873-15>.
  63. Coll F, McNeerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, Mallard K, Nair M, Miranda A, Alves A, Perdigão J, Viveiros M, Portugal I, Hasan Z, Hasan R, Glynn JR, Martin N, Pain A, Clark TG. 2015. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Med* 7:51. <https://doi.org/10.1186/s13073-015-0164-0>.
  64. Pankhurst LJ, Del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, Fermont JM, Gascoyne-Binzi DM, Kohl TA, Kong C, Lemaitre N, Niemann S, Paul J, Rogers TR, Roycroft E, Smith EG, Supply P, Tang P, Wilcox MH, Wordsworth S, Wyllie D, Xu L, Crook DW, Compass TB, Study G. 2016. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. *Lancet Respir Med* 4:49–58. [https://doi.org/10.1016/S2213-2600\(15\)00466-X](https://doi.org/10.1016/S2213-2600(15)00466-X).
  65. Sadouki Z, Day MR, Doumith M, Chattaway MA, Dallman TJ, Hopkins KL, Elson R, Woodford N, Godbole G, Jenkins C. 2017. Comparison of phenotypic and WGS-derived antimicrobial resistance profiles of *Shigella sonnei* isolated from cases of diarrhoeal disease in England and Wales, 2015. *J Antimicrob Chemother* 72:2496–2502. <https://doi.org/10.1093/jac/dkx170>.
  66. Gordon NC, Price JR, Cole K, Everitt R, Morgan M, Finney J, Kearns AM, Pichon B, Young B, Wilson DJ, Llewelyn MJ, Paul J, Peto TEA, Crook DW, Walker AS, Golubchik T. 2014. Prediction of *Staphylococcus aureus* antimicrobial resistance by whole-genome sequencing. *J Clin Microbiol* 52:1182–1191. <https://doi.org/10.1128/JCM.03117-13>.
  67. Aanensen DM, Feil EJ, Holden MTG, Dordel J, Yeats CA, Fedosejev A, Goater R, Castillo-Ramírez S, Corander J, Colijn C, Chlebowicz MA, Schouls L, Heck M, Pluister G, Ruimy R, Kahlmeter G, Åhman J, Matuschek E, Friedrich AW, Parkhill J, Bentley SD, Spratt BG, Grundmann H, European SRL Working Group. 2016. Whole-genome sequencing for routine pathogen surveillance in public health: a population snapshot of invasive *Staphylococcus aureus* in Europe. *mBio* 7:e00444-16.
  68. Kos VN, Déraspe M, McLaughlin RE, Whiteaker JD, Roy PH, Alm RA, Corbeil J, Gardner H. 2015. The resistome of *Pseudomonas aeruginosa* in relationship to phenotypic susceptibility. *Antimicrob Agents Chemother* 59:427–436. <https://doi.org/10.1128/AAC.03954-14>.
  69. Deng X, Memari N, Teatero S, Athey T, Isabel M, Mazzulli T, Fittipaldi N, Gubbay JB. 2016. Whole-genome sequencing for surveillance of invasive pneumococcal diseases in Ontario, Canada: rapid prediction of

- genotype, antibiotic resistance and characterization of emerging serotype 22F. *Front Microbiol* 7:2099. <https://doi.org/10.3389/fmicb.2016.02099>.
70. Stoesser N, Batty EM, Eyre DW, Morgan M, Wyllie DH, Del Ojo Elias C, Johnson JR, Walker AS, Peto TEA, Crook DW. 2013. Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* isolates using whole genomic sequence data. *J Antimicrob Chemother* 68:2234–2244. <https://doi.org/10.1093/jac/dkt180>.
  71. Tyson GH, McDermott PF, Li C, Chen Y, Tadesse DA, Mukherjee S, Bodeis-Jones S, Kabera C, Gaines SA, Loneragan GH, Edrington TS, Torrence M, Harhay DM, Zhao S. 2015. WGS accurately predicts antimicrobial resistance in *Escherichia coli*. *J Antimicrob Chemother* 70:2763–2769. <https://doi.org/10.1093/jac/dkv186>.
  72. Quan TP, Bawa Z, Foster D, Walker T, Del Ojo Elias C, Rathod P, MMM Informatics Group, Iqbal Z, Bradley P, Mowbray J, Walker AS, Crook DW, Wyllie DH, Peto TEA, Smith EG. 2018. Evaluation of whole-genome sequencing for mycobacterial species identification and drug susceptibility testing in a clinical setting: a large-scale prospective assessment of performance against line probe assays and phenotyping. *J Clin Microbiol* 56:e01480-17.
  73. Macedo R, Nunes A, Portugal I, Duarte S, Vieira L, Gomes JP. 2018. Dissecting whole-genome sequencing-based online tools for predicting resistance in *Mycobacterium tuberculosis*: can we use them for clinical decision guidance? *Tuberculosis (Edinb)* 110:44–51. <https://doi.org/10.1016/j.tube.2018.03.009>.
  74. Miotto P, Tessema B, Tagliani E, Chindelevitch L, Starks AM, Emerson C, Hanna D, Kim PS, Liwski R, Zignol M, Gilpin C, Niemann S, Denkinge CM, Fleming J, Warren RM, Crook D, Posey J, Gagneux S, Hoffner S, Rodrigues C, Comas I, Engelthaler DM, Murray M, Alland D, Rigouts L, Lange C, Dheda K, Hasan R, Ranganathan UDK, McNERney R, Ezewudo M, Cirillo DM, Schito M, Köser CU, Rodwell TC. 2017. A standardised method for interpreting the association between mutations and phenotypic drug resistance in *Mycobacterium tuberculosis*. *Eur Respir J* 50:1701354. <https://doi.org/10.1183/13993003.01354-2017>.
  75. Vogwill T, Kojadinovic M, MacLean RC. 2016. Epistasis between antibiotic resistance mutations and genetic background shape the fitness effect of resistance across species of *Pseudomonas*. *Proc R Soc B* 283:20160151. <https://doi.org/10.1098/rspb.2016.0151>.
  76. Trindade S, Sousa A, Xavier KB, Dionisio F, Ferreira MG, Gordo I. 2009. Positive epistasis drives the acquisition of multidrug resistance. *PLoS Genet* 5:e1000578. <https://doi.org/10.1371/journal.pgen.1000578>.
  77. Knopp M, Andersson DI. 2018. Predictable phenotypes of antibiotic resistance mutations. *mBio* 9:e00770-18.
  78. Moradigaravand D, Palm M, Farewell A, Mustonen V, Warringer J, Parts L. 2018. Precise prediction of antibiotic resistance in *Escherichia coli* from full genome sequences. *bioRxiv* <https://doi.org/10.1101/338194>.
  79. Yang Y, Niehaus KE, Walker TM, Iqbal Z, Walker AS, Wilson DJ, Peto TE, Crook DW, Smith EG, Zhu T, Clifton DA. 2017. Machine learning for classifying tuberculosis drug-resistance from DNA sequencing data. *Bioinformatics* 34:1666–1671.
  80. Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE, Jr, Walker H, Hawkins PA, Tran T, McGee L, Beall BW, Active Bacterial Core Surveillance Team. 2017. Validation of  $\beta$ -lactam minimum inhibitory concentration predictions for pneumococcal isolates with newly encountered penicillin binding protein (PBP) sequences. *BMC Genomics* 18:621. <https://doi.org/10.1186/s12864-017-4017-7>.
  81. Chen ML, Doddi A, Royer J, Freschi L, Schito M, Ezewudo M, Kohane IS, Beam A, Farhat M. 2018. Deep learning predicts tuberculosis drug resistance status from whole-genome sequencing data. *bioRxiv* <https://doi.org/10.1101/275628>.
  82. Alam MT, Petit RA, III, Crispell EK, Thornton TA, Conneely KN, Jiang Y, Satola SW, Read TD. 2014. Dissecting vancomycin-intermediate resistance in *Staphylococcus aureus* using genome-wide association. *Genome Biol Evol* 6:1174–1185. <https://doi.org/10.1093/gbe/evu092>.
  83. Davis JJ, Boisvert S, Brettin T, Kenyon RW, Mao C, Olson R, Overbeek R, Santerre J, Shukla M, Wattam AR, Will R, Xia F, Stevens R. 2016. Antimicrobial resistance prediction in PATRIC and RAST. *Sci Rep* 6:27930. <https://doi.org/10.1038/srep27930>.
  84. Nguyen M, Wesley Long S, McDermott PF, Olsen RJ, Olson R, Stevens RL, Tyson GH, Zhao S, Davis JJ. 2018. Using machine learning to predict antimicrobial minimum inhibitory concentrations and associated genomic features for nontyphoidal *Salmonella*. *bioRxiv* <https://doi.org/10.1101/380782>.
  85. Pesesky MW, Hussain T, Wallace M, Patel S, Andleeb S, Burnham C-AD, Dantas G. 2016. Evaluation of machine learning and rules-based approaches for predicting antimicrobial resistance profiles in Gram-negative bacilli from whole genome sequence data. *Front Microbiol* 7:1887.
  86. Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE, Jr, Walker H, Hawkins PA, Tran T, Whitney CG, McGee L, Beall BW. 2016. Penicillin-binding protein transpeptidase signatures for tracking and predicting  $\beta$ -lactam resistance levels in *Streptococcus pneumoniae*. *mBio* 7:e00756-16.
  87. Metcalf BJ, Gertz RE, Jr, Gladstone RA, Walker H, Sherwood LK, Jackson D, Li Z, Law C, Hawkins PA, Chochua S, Sheth M, Rayamajhi N, Bentley SD, Kim L, Whitney CG, McGee L, Beall B, Active Bacterial Core surveillance team. 2016. Strain features and distributions in pneumococci from children with invasive disease before and after 13-valent conjugate vaccine implementation in the USA. *Clin Microbiol Infect* 22:60.e9–60.e29. <https://doi.org/10.1016/j.cmi.2015.08.027>.
  88. Vogwill T, MacLean RC. 2015. The genetic basis of the fitness costs of antimicrobial resistance: a meta-analysis approach. *Evol Appl* 8:284–295. <https://doi.org/10.1111/eva.12202>.
  89. Wadsworth CB, Arnold BJ, Sater MRA, Grad YH. 2018. Azithromycin resistance through interspecific acquisition of an epistasis dependent efflux pump component and transcriptional regulator in *Neisseria gonorrhoeae*. *bioRxiv* <https://doi.org/10.1101/309294>.
  90. Skwark MJ, Croucher NJ, Puranen S, Chewapreecha C, Pesonen M, Xu YY, Turner P, Harris SR, Beres SB, Musser JM, Parkhill J, Bentley SD, Aurell E, Corander J. 2017. Interacting networks of resistance, virulence and core machinery genes identified by genome-wide epistasis analysis. *PLoS Genet* 13:e1006508. <https://doi.org/10.1371/journal.pgen.1006508>.
  91. Guérillot R, Li L, Baines S, Howden BO, Schultz M, Seemant T, Monk I, Pidot SJ, Gao W, Giulieri S, da Silva AG, D'Agata A, Tomita T, Peleg AY, Stinear TP, Howden BP. 2018. Comprehensive antibiotic-linked mutation assessment by resistance mutation sequencing (RM-seq). *bioRxiv* <https://doi.org/10.1101/257915>.
  92. Guérillot R, da Silva AG, Monk I, Giulieri S, Tomita T, Alison E, Porter J, Pidot S, Gao W, Peleg AY, Seemann T, Stinear TP, Howden BP. 2018. Convergent evolution driven by rifampin exacerbates the global burden of drug-resistant *Staphylococcus aureus*. *mSphere* 3:e00550-17.
  93. Wistrand-Yuen E, Knopp M, Hjort K, Koskiniemi S, Berg OG, Andersson DI. 2018. Evolution of high-level resistance during low-level antibiotic exposure. *Nat Commun* 9:1599. <https://doi.org/10.1038/s41467-018-04059-1>.
  94. McEvoy CRE, Tsuji B, Gao W, Seemann T, Porter JL, Doig K, Ngo D, Howden BP, Stinear TP. 2013. Decreased vancomycin susceptibility in *Staphylococcus aureus* caused by IS256 tempering of WalkR expression. *Antimicrob Agents Chemother* 57:3240–3249. <https://doi.org/10.1128/AAC.00279-13>.
  95. Di Gregorio S, Fernandez S, Perazzi B, Bello N, Famiglietti A, Mollerach M. 2016. Increase in IS256 transposition in invasive vancomycin heteroresistant *Staphylococcus aureus* isolate belonging to ST100 and its derived VISA mutants. *Infect Genet Evol* 43:197–202. <https://doi.org/10.1016/j.meegid.2016.05.001>.
  96. Wheeler NE, Barquist L, Kingsley RA, Gardner PP. 2016. A profile-based method for identifying functional divergence of orthologous genes in bacterial genomes. *Bioinformatics* 32:3566–3574. <https://doi.org/10.1093/bioinformatics/btw518>.
  97. Fowler PW, Cole K, Claire Gordon N, Kearns AM, Llewelyn MJ, Peto TEA, Crook DW, Sarah Walker A. 2018. Robust prediction of resistance to trimethoprim in *Staphylococcus aureus*. *Cell Chem Biol* 25:339.e4–349.e4.
  98. Zhang Y, Mitchison D. 2003. The curious characteristics of pyrazinamide: a review. *Int J Tuberc Lung Dis* 7:6–21.
  99. Van Deun A, Barrera L, Bastian I, Fattorini L, Hoffmann H, Kam KM, Rigouts L, Rüsç-Gerdes S, Wright A. 2009. *Mycobacterium tuberculosis* strains with highly discordant rifampin susceptibility test results. *J Clin Microbiol* 47:3501–3506. <https://doi.org/10.1128/JCM.01209-09>.
  100. Van Deun A, Aung KJM, Bola V, Lebeke R, Hossain MA, de Rijk WB, Rigouts L, Gumusboga A, Torrea G, de Jong BC. 2013. Rifampin drug resistance tests for tuberculosis: challenging the gold standard. *J Clin Microbiol* 51:2633–2640. <https://doi.org/10.1128/JCM.00553-13>.
  101. Horne DJ, Pinto LM, Arentz M, Lin S-YG, Desmond E, Flores LL, Steingart KR, Minion J. 2013. Diagnostic accuracy and reproducibility of WHO-endorsed phenotypic drug susceptibility testing methods for first-line and second-line antituberculosis drugs. *J Clin Microbiol* 51:393–401. <https://doi.org/10.1128/JCM.02724-12>.
  102. Beaudoin T, Yau YCW, Stapleton PJ, Gong Y, Wang PW, Guttman DS, Waters V. 2017. *Staphylococcus aureus* interaction with *Pseudomonas*

- aeruginosa biofilm enhances tobramycin resistance. NPJ Biofilms Microbiomes 3:25. <https://doi.org/10.1038/s41522-017-0035-0>.
103. Radlinski L, Rowe SE, Kartchner LB, Maile R, Cairns BA, Vitko NP, Gode CJ, Lachiewicz AM, Wolfgang MC, Conlon BP. 2017. Pseudomonas aeruginosa exoproducts determine antibiotic efficacy against Staphylococcus aureus. PLoS Biol 15:e2003981. <https://doi.org/10.1371/journal.pbio.2003981>.
  104. Donlan RM, Costerton JW. 2002. Biofilms: survival mechanisms of clinically relevant microorganisms. Clin Microbiol Rev 15:167–193. <https://doi.org/10.1128/CMR.15.2.167-193.2002>.
  105. Cornforth DM, Dees JL, Ibberson CB, Huse HK, Mathiesen IH, Kirketerp-Møller K, Wolcott RD, Rumbaugh KP, Bjarnsholt T, Whiteley M. 2018. Pseudomonas aeruginosa transcriptome during human infection. Proc Natl Acad Sci U S A 115:E5125–E5134. <https://doi.org/10.1073/pnas.1717525115>.
  106. Hassan MH, Butler MS, Ranzoni A, Cooper MA. 2018. Detection and quantification of the heterogeneity of S. aureus bacterial populations to identify antibiotic-induced persistence. bioRxiv <https://doi.org/10.1101/320093>.
  107. Kassim A, Omuse G, Premji Z, Revathi G. 2016. Comparison of Clinical Laboratory Standards Institute and European Committee on Antimicrobial Susceptibility Testing guidelines for the interpretation of antibiotic susceptibility at a University teaching hospital in Nairobi, Kenya: a cross-sectional study. Ann Clin Microbiol Antimicrob 15:21. <https://doi.org/10.1186/s12941-016-0135-3>.
  108. Pou C, Noguera-Julian M, Pérez-Álvarez S, García F, Delgado R, Dalmau D, Álvarez-Tejado M, Gonzalez D, Sayada C, Chueca N, Pulido F, Ibáñez L, Rodríguez C, Casadellà M, Santos JR, Ruiz L, Clotet B, Paredes R. 2014. Improved prediction of salvage antiretroviral therapy outcomes using ultrasensitive HIV-1 drug resistance testing. Clin Infect Dis 59:578–588. <https://doi.org/10.1093/cid/ciu287>.
  109. Codoñer FM, Pou C, Thielen A, García F, Delgado R, Dalmau D, Álvarez-Tejado M, Ruiz L, Clotet B, Paredes R. 2011. Added value of deep sequencing relative to population sequencing in heavily pre-treated HIV-1-infected subjects. PLoS One 6:e19461. <https://doi.org/10.1371/journal.pone.0019461>.
  110. Runtuwene LR, Tuda JSB, Mongan AE, Makalowski W, Frith MC, Imwong M, Srisutham S, Nguyen Thi LA, Tuan NN, Eshita Y, Maeda R, Yamagishi J, Suzuki Y. 2018. Nanopore sequencing of drug-resistance-associated genes in malaria parasites, Plasmodium falciparum. Sci Rep 8:8286. <https://doi.org/10.1038/s41598-018-26334-3>.
  111. Crisan A, McKee G, Munzner T, Gardy JL. 2018. Evidence-based design and evaluation of a whole genome sequencing clinical report for the reference microbiology laboratory. PeerJ 6:e4218. <https://doi.org/10.7717/peerj.4218>.
  112. Gargis AS, Kalman L, Lubin IM. 2016. Assuring the quality of next-generation sequencing in clinical microbiology and public health laboratories. J Clin Microbiol 54:2857–2865. <https://doi.org/10.1128/JCM.00949-16>.
  113. Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8:186–194. <https://doi.org/10.1101/gr.8.3.186>.
  114. Liu J, Chen D, Peters BM, Li L, Li B, Xu Z, Shirliff ME. 2016. Staphylococcal chromosomal cassettes mec (SCCmec): a mobile genetic element in methicillin-resistant Staphylococcus aureus. Microb Pathog 101:56–67. <https://doi.org/10.1016/j.micpath.2016.10.028>.
  115. Zankari E, Hasman H, Kaas RS, Seyfarth AM, Agersø Y, Lund O, Larsen MV, Aarestrup FM. 2013. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. J Antimicrob Chemother 68:771–777. <https://doi.org/10.1093/jac/dks496>.
  116. Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, Bradley P, Iqbal Z, Feuerriegel S, Niehaus KE, Wilson DJ, Clifton DA, Kapatai G, Ip CLC, Bowden R, Drobniewski FA, Allix-Béguec C, Gaudin C, Parkhill J, Diel R, Supply P, Crook DW, Smith EG, Walker AS, Ismail N, Niemann S, Peto TEA, Modernizing Medical Microbiology (MMM) Informatics Group. 2015. Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. Lancet Infect Dis 15:1193–1202. [https://doi.org/10.1016/S1473-3099\(15\)00062-6](https://doi.org/10.1016/S1473-3099(15)00062-6).
  117. Nguyen M, Brettin T, Long SW, Musser JM, Olsen RJ, Olson R, Shukla M, Stevens RL, Xia F, Yoo H, Davis JJ. 2018. Developing an in silico minimum inhibitory concentration panel test for Klebsiella pneumoniae. Sci Rep 8:421.
  118. Long SW, Olsen RJ, Eagar TN, Beres SB, Zhao P, Davis JJ, Brettin T, Xia F, Musser JM. 2017. Population genomic analysis of 1,777 extended-spectrum beta-lactamase-producing Klebsiella pneumoniae isolates, Houston, Texas: unexpected abundance of clonal group 307. mBio 8:e00489-17.