

# Genomic and Metagenomic Approaches for Predictive Surveillance of Emerging Pathogens and Antibiotic Resistance

Kimberley V. Sukhum<sup>1,2,†</sup>, Luke Diorio-Toth<sup>1,†</sup> and Gautam Dantas<sup>1,2,3,4,\*</sup>

Antibiotic-resistant organisms (AROs) are a major concern to public health worldwide. While antibiotics have been naturally produced by environmental bacteria for millions of years, modern widespread use of antibiotics has enriched resistance mechanisms in human-impacted bacterial environments. Antibiotic resistance genes (ARGs) continue to emerge and spread rapidly. To combat the global threat of antibiotic resistance, researchers must develop methods to rapidly characterize AROs and ARGs, monitor their spread across space and time, and identify novel ARGs and resistance pathways. We review how high-throughput sequencing-based methods can be combined with classic culture-based assays to characterize, monitor, and track AROs and ARGs. Then, we evaluate genomic and metagenomic methods for identifying ARGs and biosynthetic pathways for novel antibiotics from genomic data sets. Together, these genomic analyses can improve surveillance and prediction of emerging resistance threats and accelerate the development of new antibiotic therapies to combat resistance.

Since Paul Ehrlich's discovery that arsphenamine could cure syphilis-infected rabbits and Alexander Fleming's serendipitous discovery of the natural product penicillin, antibiotics have been among the most successful chemotherapies in the history of medicine.<sup>1</sup> In addition to directly treating infectious diseases, a leading cause of morbidity and mortality for most of human history, antibiotics enable treatments like invasive surgery, organ transplantation, and cancer chemotherapy, which would otherwise be unsafe due to the high risk of infection.<sup>2</sup> The emergence of widespread antibiotic resistance significantly compromises these advancements. The World Health Organization (WHO) described antibiotic resistance as the single greatest challenge in infectious disease today, threatening high and low gross domestic product countries alike.<sup>3</sup> One analysis of data from the European Union estimated that in 2015 there were over 600,000 infections and 33,000 deaths attributable to antibiotic-resistant organisms (AROs), figures that have increased by nearly 2.5-fold compared with similar estimates from 2007.<sup>4</sup>

Although human use of antibiotics began in the 20th century, antibiotics have been naturally produced by environmental bacteria for millions of years.<sup>3</sup> Because antibiotic production appears to be a natural feature of bacterial communities, it should not be surprising that defense mechanisms against those compounds appear as an evolutionary consequence.<sup>5</sup> Sequencing of DNA preserved in ancient environmental samples has detected diverse antibiotic resistance genes (ARGs) in metagenomes of 30,000-year-old permafrost sediments and bacteria recovered from a cave isolated

from human contact for 4 million years.<sup>6,7</sup> Similarly, sequencing of the host-associated microbiomes of previously uncontacted Amerindians detected numerous ARGs against modern antibiotics.<sup>8</sup> The collection of resistance genes in a given microbiome is known as the *resistome* (Table 1). Worryingly, longitudinal analyses of resistomes have suggested that human use of antibiotics have influenced their composition.<sup>3,9</sup>

Selective pressure from widespread use of antibiotics in agricultural and clinical settings has enriched natural resistance mechanisms in human-impacted bacterial environments. Analysis of soil samples from 1940 to the present day revealed increased abundance of ARGs to a wide range of antibiotics, and clinical resistance to every class of antibiotics has been observed, sometimes within the same year as deployment.<sup>3,9</sup> This is further highlighted by the rapid global spread of plasmid-borne resistance determinants like *mcr1*, *bla*<sub>KPC</sub>, *bla*<sub>NDM-1</sub>, and *qnr*, which provide resistance to antibiotics of last resort.<sup>5</sup> However, the clinical focus on antibiotic resistance in human pathogens has historically led to an undersampling and underestimate of the diverse array of resistance determinants that exist in environmental bacteria, which may be exchanged with pathogenic hosts through horizontal gene transfer.<sup>10–13</sup> Therefore, the study of antibiotic resistance has recently expanded from examination of isolated, cultivatable, and easily cultured pathogens to include metagenomic approaches that capture resistomes of microbial communities across diverse habitats.<sup>10,14,15</sup> These genomic assays are ideally combined with functional and culture-based assays

<sup>1</sup>The Edison Family Center for Genome Sciences and Systems Biology, Washington University in St. Louis School of Medicine, St. Louis, Missouri, USA;

<sup>2</sup>Department of Pathology and Immunology, Washington University in St. Louis School of Medicine, St. Louis, Missouri, USA; <sup>3</sup>Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, Missouri, USA; <sup>4</sup>Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, Missouri, USA. \*Correspondence: Gautam Dantas ([dantas@wustl.edu](mailto:dantas@wustl.edu))

<sup>†</sup>These authors contributed equally to this work.

Received March 20, 2019; accepted May 14, 2019. doi:10.1002/cpt.1535

**Table 1 Glossary of terms**

Term	Definition
Antibiotic resistance gene (ARG)	A gene that confers an organism the ability to resist an antibiotic treatment that would otherwise affect it
Antibiotic-resistant organism (ARO)	An organism that is not controlled or killed by an antibiotic treatment that would otherwise affect it
Biosynthetic gene cluster (BGC)	A physically clustered group of two or more genes in a particular genome that together encode a biosynthetic pathway for the production of a specialized metabolite
Commensal	A microbe that coexists with a host without detriment or obvious benefit to the host
Contig	Contiguous DNA sequence assembled from shorter sequencing reads
Homology	In the context of genomics, the existence of a shared ancestry between two genes
Horizontal gene transfer	The transfer of genes between microbes that are not in a parent–offspring relationship. Usually occurs through transformation, conjugation, or transduction
Isolate	A pure strain of microbe that has been separated from a mixed microbial culture
Microbiome	The collective community of microbial genomes in a given environment or sample
Next-generation sequencing (NGS)	A collection of DNA sequencing technologies that allow for rapid generation of sequencing data that is orders of magnitude larger in volume than Sanger sequencing. Includes technologies like Illumina and Ion Torrent sequencing
Open reading frame (ORF)	A sequence of DNA that is translated into protein
Pathogen	An infectious agent whose presence results in disease or damage to the host
Profile Hidden Markov model (pHMM)	A statistical model where hidden states emit observed sequences with a defined probability. Widely used for genome annotation, where protein/gene families are modeled as “states” and the primary amino acid/nucleotide sequence is the observed sequence
Resistome	The collection of resistance genes in a given environment or sample. This includes intrinsic, acquired, proto-resistance, and cryptic resistance genes.
Proto-resistance gene	A gene that has the potential to evolve into a resistance element
Cryptic resistance gene	A resistance gene that is embedded in the bacterial chromosome but does not obviously confer resistance due to low expression
Strain	A genetic variant or subtype of a microbial species
Whole genome sequencing (WGS)	The complete sequencing of genes from an organism
Whole metagenome sequencing (WMS)	The complete sequencing of genes from all organisms in a given environment or sample

Illumina (San Diego, CA); Ion Torrent (Thermo Fisher Scientific, Waltham, MA).

to improve surveillance and prediction of emerging resistance threats.<sup>16,17</sup> Further, by characterizing the vast biosynthetic capacity of the microbiome, we can accelerate the development of new antibiotic therapies to combat resistance.<sup>18,19</sup>

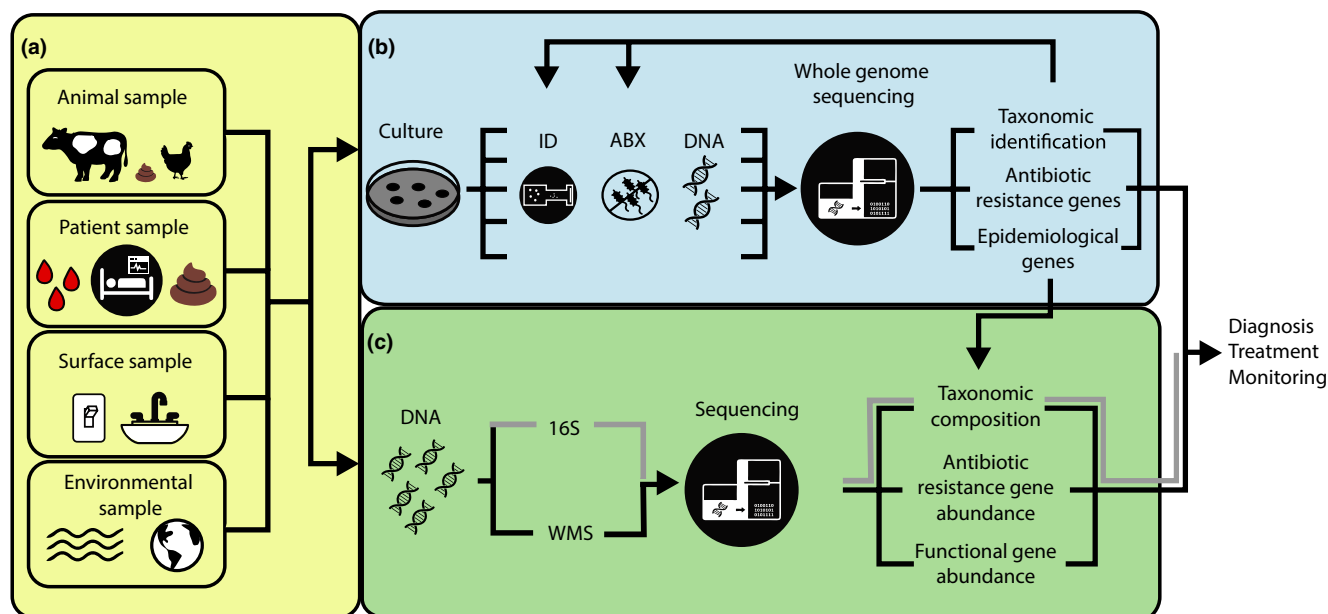
In this review, we provide an overview of how high-throughput sequencing-based methods can be combined with classic culture-based assays to improve diagnosis and treatment of infectious disease. We examine how these new methods are used to characterize, monitor, and track resistance determinants. Next, we critically evaluate the genomic and metagenomic methods that have been developed for identifying ARGs and biosynthetic pathways for novel antibiotics from genomic data sets, and highlight exciting new methods that have utilized sequencing technologies to explore genomic diversity beyond what nature can provide.

#### USING GENOMICS TO BETTER CHARACTERIZE AND IDENTIFY AROS AND ARGs

Surveying the presence and abundance of AROs is frequently done through culture-based methods, where microbial colonies are isolated and identified by phenotypic characteristics, such as morphology, size, and pigmentation, or by chemical profile using matrix-assisted laser desorption/ionization time-of-flight

mass spectrometry (MALDI-TOF) (**Figure 1a,b**).<sup>20</sup> After identification, phenotypic antibiotic susceptibility of isolates can be determined by disk diffusion, antibiotic gradient, or broth micro-dilution to determine minimum inhibitory concentrations (**Figure 1b**).<sup>21</sup> Together, culture-based methods allow for characterization of isolates but are limited in speed and scale.<sup>22</sup> Because of this, screening a large number and variety of isolates from both patients and environmental settings can be a tedious and expensive process.<sup>22</sup> Recent studies in research and diagnostic microbiology pair culture-based methods with next-generation sequencing (NGS; **Figure 1b**). Advances in NGS throughput have enabled relatively inexpensive and fast whole genome sequencing (WGS) of isolates, permitting WGS to be used more readily in both research and diagnostic microbiology.<sup>23</sup> WGS analyses allow for accurate identification of isolate strain phylogeny, ARGs, and epidemiological traits, which is critical when tracking outbreaks<sup>16,24</sup> (**Figure 1b**).

The future of monitoring and combating the presence of AROs and ARGs lies in successfully integrating and optimizing culture-based methods with WGS analyses to rapidly identify, characterize, and track emerging pathogens. Diverse microbial species and strains can have very different pathological patterns, and accurate



**Figure 1** Biological and environmental samples can be interrogated for antibiotic-resistant organism (ARO) and antibiotic resistance gene (ARG) enrichment using whole genome sequencing (WGS) and metagenomics analyses. (a) Samples can be collected from diverse origins, including biological samples such as tissue, blood, or fecal samples from animal or patient samples. Environmental samples can be collected directly from built environment surfaces or natural environments, such as soil and water sources. (b) In culture-based methods, isolates are extracted from samples and processed individually to determine species identification and antibiotic susceptibility. After DNA extraction, each isolate can be sequenced using WGS. WGS analyses can determine taxonomic identification, antibiotic resistance genotype, and epidemiological genes. These characteristics can be compared with culture-based results, creating genotype-phenotype characterization of isolates and improved characterization of species and resistome. WGS analyses can also be used to improve sequence-based searches for metagenomics analyses. (c) In metagenomics analyses, genomic DNA is extracted directly from a sample, then 16S or whole metagenomics sequencing (WMS) is done. After 16S amplification and sequencing, taxonomic composition is determined. After WMS, taxonomic composition, antibiotic resistance gene abundance, and functional gene abundance can be analyzed. All of these data can be combined to improve diagnosis, treatment, and ARO and ARG surveillance. ABX, antibiotics; ID, identification.

identification of strain and characterization of resistance profiles is critical when monitoring and treating pathogens in humans.<sup>24,25</sup> For example, *Klebsiella* is a genus of gram-negative bacteria that is commonly found in both environmental and human-associated samples. *Klebsiella pneumoniae* is well known for being an opportunistic human pathogen associated with nosocomial infections and worldwide antibiotic resistance.<sup>26</sup> The related species *Klebsiella variicola*, identified in 2003, was considered less pathogenic; however, recent studies suggest that 2–10% of *K. pneumoniae* strains isolated from infections are actually misidentifications of *K. variicola*, indicating that *K. variicola* is more prevalent in infections than initially thought.<sup>26–28</sup> Further, there is an increased mortality in patients infected with *K. variicola* compared with patients with *K. pneumoniae*, and members of the two species have different resistance profiles.<sup>24,27,28</sup> By integrating culture-based and WGS analyses, a recent study successfully reclassified commonly misidentified *K. variicola* isolates and better characterized phylogeny and resistance profiles of this emerging pathogen.<sup>24</sup> This study uncovered many high-risk ARGs in *K. variicola* genomes from a variety of geographic locations.<sup>24</sup> Further, by using WGS to investigate key virulence factors, this study demonstrated that genomic changes in *K. variicola* pilus genes were associated with uropathogenicity in a mouse model of urinary tract infections.<sup>24</sup> Altogether, recent studies of *K. variicola* highlight the importance of accurate identification and characterization of emerging pathogens and

how culture-based and WGS analyses can be integrated to gain a better understanding of emerging pathogens.<sup>24,26–28</sup>

#### USING GENOMICS TO BETTER TRACK OUTBREAKS AND MONITOR RESERVOIRS OF AROS AND ARGs

Beyond identification and characterization of AROs, it is critical to track and predict antibiotic-resistant outbreaks and reservoirs of AROs and ARGs.<sup>16,17,29</sup> To do this, studies must be able to accurately identify the relatedness and antibiotic resistance profiles of isolates. A recent study in Australia used WGS and culture-based methods to track an emerging pathogen, *vanA*-possessing vancomycin-resistant *Enterococcus faecium* (VREfm).<sup>16</sup> Prior to 2015, *vanA* VREfm was rare in Australia but has subsequently had a dramatic increase with 44% of the VREfm isolates identified as *vanA* in 2016.<sup>16</sup> WGS of VREfm isolates was used to determine the phylogenetic relationship of this emerging pathogen in Australia. Using core genome single nucleotide polymorphism diversity, this study found 26 of the isolates from 24 patients in the outbreak were from the same strain, 17 of those isolates could not be differentiated at the core genome level, and the remaining 9 isolates differed from the main cluster by 1–6 single nucleotide polymorphisms.<sup>16</sup> This resolution allowed the study to track the emergence of colonization and clinical infection from *vanA* VREfm.<sup>16</sup> In another study, clonal and strain tracking was used to determine the spread of a multidrug-resistant *Sphingomonas koreensis* outbreak

in the United States.<sup>17</sup> *Sphingomonas* species are common in built and aquatic environments but have been implicated in a range of infections, including pneumonia, meningitis, catheter-associated bloodstream infections, and wound infections.<sup>17</sup> This study investigated six infections of *Sphingomonas* over a 6-month period at the US National Institutes of Health (NIH). Core genome single nucleotide variants were used to identify and track strains across patients and clinical water sources.<sup>17</sup> Four of the *S. koreensis* patient isolates had high genomic similarity to each other (>99.9% average nucleotide identity (ANI)) and to environmental *S. koreensis* isolates from water and faucet samples collected from inside patient rooms (>99.7% ANI).<sup>17</sup> This genomic similarity suggested that there was a common environmental reservoir for this pathogen that had persisted in the hospital infrastructure over time.<sup>17</sup> Further, banked *S. koreensis* isolates from 2006 also showed high similarity to isolates from the 2016 outbreak (>99.8% ANI), suggesting this reservoir had persisted for a decade.<sup>17</sup> These studies demonstrate the capabilities of WGS in tracking individual strains in outbreaks.

Beyond tracking the strains across an outbreak, WGS analyses can be used to determine the effectiveness of mediation strategies to prevent future ARO outbreaks. One such study tracked a large *K. pneumoniae* carbapenemase (KPC)-producing *Escherichia coli* outbreak to determine the impact of ARO decontamination intervention after an outbreak. KPC-producing *E. coli* are uncommon, and the emergence of an outbreak with carbapenem resistance is a particular concern because of the extremely limited treatment options for such infections.<sup>29</sup> In 2015, a sudden increase of KPC-producing *E. coli* colonization was detected in patient fecal samples at the Central Manchester University Hospital NHS Foundation Trust (United Kingdom), spurring a genomic epidemiological study on isolates from patients and the environment. While environmental samples were taken from sinks, drains, toilets, sluices, and other high-touch sites, all carbapenem-resistant Enterobacteriaceae cultures were associated with wastewater sites.<sup>29</sup> WGS analyses indicated that the outbreak was due to the spread of a KPC-producing *E. coli* clone among both patients and the environment.<sup>29</sup> A decontamination intervention was performed, and carbapenem-resistant Enterobacteriaceae incidence declined after the wards were closed and the plumbing replaced.<sup>29</sup> However, shortly after the intervention, wastewater sites were rapidly recolonized again by the KPC-producing clone, suggesting the importance of developing better containment and elimination strategies.<sup>29</sup> Together, these studies demonstrate how WGS can be used as a tool to better characterize emerging pathogens, monitor reservoirs of AROs and ARGs, and evaluate success of reservoir decontamination interventions within hospitals.

While WGS continues to become quicker and less expensive, the technology is more expensive and computationally intensive than other clinical methods to identify and characterize pathogens.<sup>30</sup> Sequencing costs can vary from \$100 to \$500 a genome based on a number of factors, including type of sequencing, genome size, depth of sequencing, and number of samples processed.<sup>31,32</sup> Further, a maintained bioinformatics pipeline is essential to store, process, and interpret WGS for clinical relevance. This pipeline comes with associated costs and additional necessary expertise.

Because of these costs, WGS currently does not replace the need for other clinical methods for individual patient diagnosis and is best applied when characterizing emerging pathogens or outbreaks of AROs.

## USING METAGENOMICS TO IMPROVE IDENTIFICATION OF AROS AND CLINICAL TREATMENT

Culture-based methods provide valuable insights to ARO colonization and allow for WGS investigation into ARO strains across outbreaks. However, these methods often restrict studies to a few genes of interest or to easily cultured organisms, precluding studies on the emergence and spread of AROs and ARGs in diverse and complex uncultured microbial communities. As ARGs can spread from environmental organisms to human pathogens, it is essential to survey and characterize environmental microbes alongside human pathogens.<sup>10</sup> Uncultured microbial communities can be interrogated using metagenomics, wherein genomic DNA is extracted directly from a sample without the need to first culture microorganisms (Figure 1c). After DNA extraction, there are two prominent methods to characterize whole microbial communities: 16S sequencing and whole metagenome sequencing (WMS). 16S metagenomic studies amplify and sequence a small variable segment of the 16S RNA gene. The 16S rRNA gene is highly conserved and present in almost all bacteria. From 16S sequences, it is possible to infer microbial taxonomic presence and relative abundance from a sample. In WMS, all genomic DNA is fragmented and sequenced directly after extraction without amplification. By comparing the fragmented shotgun reads with databases of known functional and resistance genes, it is possible to determine relative abundance of taxa and known functional and resistance genes<sup>33–35</sup> (Figure 1c). Both metagenomic methods are relatively quick and high throughput compared with culture-based methods.<sup>3</sup> Furthermore, longitudinal metagenomic approaches are well suited for addressing questions of burden and accumulation of antibiotic resistance. Samples taken from the same patient or population across treatment will allow researchers to directly address how resistance has changed and determine strategies to minimize the emergence and transmission of resistance. Thus, metagenomic analyses can be a useful tool to monitor and characterize AROs and ARGs in areas where culture-based methods are insufficient or inconvenient.

While culture-based methods are the standard by which to diagnose and confirm infectious agents in clinical microbiology, many samples entering clinical laboratories return negative despite being subjected to a battery of conventional laboratory tests.<sup>25,36</sup> In cases where a causative agent of infection is not identified, metagenomic analyses can be used to rapidly identify potential pathogens.<sup>25</sup> For example, hundreds of pathogens have been associated with infectious encephalitis, and in as many as 60% of encephalitis cases, conventional laboratory assays fail to detect a causative agent.<sup>37</sup> Accurate and rapid identification of the infectious agent is critical as the main alternative diagnosis is immune-mediated encephalitis, which requires immune suppression for treatment.<sup>37</sup> In one striking case, an HIV-infected patient was admitted to a hospital with lethally diffuse brain lesions.<sup>25</sup> The patient's blood and cerebrospinal fluid tests were negative for common microbial infections.



Unbiased pathogen detection by metagenomic sequencing of cerebrospinal fluid found sequences mapping to *Toxoplasma gondii*. After identification, the patient was immediately treated with azithromycin plus co-trimoxazole for toxoplasmosis encephalitis and showed significant short-term improvement, but for complex reasons was lost to follow-up.<sup>25</sup> *T. gondii*, while a known neurological opportunistic pathogen, is difficult to culture and not easily detected by microscopic examination. Polymerase chain reaction (PCR) assays for detection typically require an *a priori* suspicion of *T. gondii* and are not readily available in many regions.<sup>25</sup> In this case and many others like it, WMS analyses have the capability to directly and rapidly identify pathogen presence, improving diagnosis and treatment of patients.

Metagenomics can also be used to rapidly monitor dysbiotic states of the microbiome that may induce or be correlated with disease and disease progression. For example, early detection and intervention is critical for treating liver cirrhosis.<sup>38</sup> Early interventions can avoid or delay patient decompensation, but chronic liver disease is often asymptomatic in early cases, and diagnosis for cirrhosis is typically done after liver biopsy, an invasive procedure that is not applicable to all patients and can lead to complications of its own.<sup>38</sup> However, metagenomic studies have found a direct relationship between gut dysbiosis and the progression of cirrhosis in patients.<sup>38</sup> In these cases, metagenomic analyses can be used to determine the progression of liver cirrhosis by monitoring the depletion of species and gene richness in stool samples. This has the potential to improve detection and treatment across the course of disease progression.<sup>38</sup>

There are also limitations to metagenomic analyses from clinical samples that must be considered when evaluating metagenomic analyses as a tool in clinical diagnosis. While metagenomic analyses may have higher sensitivity in pathogen detection and identification than culture-based methods,<sup>25,36</sup> it has yet to be determined if metagenomic analyses have higher sensitivity than other molecular-based diagnosis methods, such as pathogen-specific PCR.<sup>25</sup> Molecular-based methods have been used successfully and rapidly in many clinical cases; however, their clinical sensitivity can be highly variable and they are more prone to false positives due to contamination or non-causative agents.<sup>39</sup> Further, in human samples, the amount of human genomic DNA often vastly outnumbers the amount of microbial genomic DNA.<sup>40</sup> This disparity can greatly influence the read depth necessary to detect microbial presence. Tools are being developed that selectively deplete eukaryotic DNA, allowing enhanced interrogation of microbial presence.<sup>40</sup> Metagenomics also shares the same sequencing and computation costs discussed for WGS. Because of these costs and limitations, metagenomics is applied in patients after an infectious agent cannot be identified by other means and in human samples where there is a known microbial presence.

#### USING METAGENOMICS TO MONITOR ENVIRONMENTAL AND BIOLOGICAL RESERVOIRS OF ANTIBIOTIC RESISTANCE

The spread of antibiotic-resistant infectious agents is of rising global concern. One specific example of this is *bla*<sub>NDM-1</sub>, which can confer resistance to a wide range of  $\beta$ -lactam antibiotics, including carbapenems, a class of last-resort antibiotics.<sup>41</sup> *bla*<sub>NDM-1</sub>

was initially discovered in 2008 with a suspected origin of India.<sup>41</sup> Since then, different bacterial hosts, such as *Klebsiella*, and *E. coli*, etc., containing *bla*<sub>NDM-1</sub> have been associated with outbreaks in hospitals across a wide geographic area, including all continents except South America and Antarctica.<sup>41</sup> Furthermore, *bla*<sub>NDM-1</sub> has been identified in drinking water samples in the Indian subcontinent.<sup>42</sup> While *bla*<sub>NDM-1</sub> is not unique in its clinical relevance, it stands as a concerning example of how ARGs can disseminate rapidly in the environment to compromise disease treatment in humans. A metagenomic characterization study found *bla*<sub>NDM-1</sub> in hospital wastewater, indicating wastewaters as a hotspot for ARGs.<sup>43</sup> An important aspect of understanding and combating the rising prevalence of antibiotic resistance is to survey and monitor environmental reservoirs for infectious agents and accumulation of ARGs. The human microbiome is regularly exposed and influenced by environmental bacteria.<sup>44,45</sup> These environmental bacteria could act as a reservoir of ARGs, which may be exchanged with pathogens that in turn could infect humans.<sup>10</sup> To best predict the emergence and abundance of AROs and ARGs, researchers must characterize and predict the spread of environment AROs and ARGs and their influence on human pathogens. While many environments can act as reservoirs of infectious agents, host-associated commensals and water reservoirs have recently been highlighted as particularly important to monitor.<sup>13,46–48</sup> Metagenomic analyses allow for long-term and rapid reservoir monitoring that can then be directed and verified using culture-based methods.

Domesticated animals pose a potentially dangerous biological reservoir for AROs and ARGs because of the high risk and relative ease of zoonotic transmission to humans.<sup>49,50</sup> Antibiotics at both therapeutic and nontherapeutic concentrations are used extensively in livestock (e.g., poultry, cattle, swine), with total annual volumes far exceeding human antibiotic use.<sup>51</sup> Remnant antibiotic compounds and their metabolites are excreted in livestock manure, litter, and wastewater.<sup>47,51,52</sup> Metagenomic surveillance of animal waste products shows an enrichment of ARGs in environmental bacteria associated with these animal reservoirs.<sup>51,52</sup> In 2015, the emergence of the first plasmid-mediated colistin resistance gene, *mcr-1*, was found from a surveillance project of commensal *E. coli* in pigs in China.<sup>13</sup> Colistin is a last-resort antibiotic, and before the discovery of *mcr-1*, colistin resistance had only been observed through chromosomal mutations.<sup>13</sup> A plasmid-mediated resistance gene allows the gene to be easily maintained in populations and transfer rapidly into pathogenic strains.<sup>13</sup> Since the discovery of *mcr-1*, it has been found in animal and food sources and in community-acquired infections.<sup>53</sup> Monitoring ARG enrichment in animal reservoirs may help predict emerging human pathogens. However, the magnitude by which AROs or ARGs in animal reservoirs contribute to human pathogens and infections remains unclear, and further studies are required to provide quantitative assessments of these risks.

Water supplies are home to a wide diversity of microbes and can become contaminated by known human pathogens that have been categorized as serious or urgent clinical antimicrobial resistant threats.<sup>29,46,48</sup> Given that many acquired ARGs are harbored on conjugative plasmids that can self-mobilize between diverse bacterial species, water environments represent a reservoir where ARGs

can exchange between benign bacteria and human pathogens.<sup>48</sup> Wastewaters receive diverse discharges with a high presence and abundance of microbial organisms.<sup>54,55</sup> These discharges include antibiotic remnants from use in humans and animals, which may increase selection pressures on ARGs native to water microbial communities (**Figure 2**). Although wastewater is treated to eliminate coliform bacteria, treatments may be ineffective at destroying ARGs, resulting in ARG enrichment in surviving wastewater microbes.<sup>55–57</sup> Further, functional metagenomic studies (methods described below) have identified increases in phenotypically resistant clones recovered from samples downstream of wastewater treatment plants, suggesting that wastewater treatment and disposal increases the reservoir of resistance mechanisms in the environment.<sup>58</sup> One recent study used metagenomics to better understand ARO and ARG abundance in the Chaobia River, a major source of drinking water, agricultural irrigation, and waste in China.<sup>59</sup> Metagenomic analyses demonstrated that the Chaobia River has elevated ARG abundance compared with other areas in the region and with areas without human interaction such as deep-sea and Antarctic sediment. The dominant ARGs found were associated with antibiotics that had been heavily used in human medicine, veterinary therapy, and livestock growth.<sup>59</sup> This increase in ARGs in major water sources may have a direct impact on millions of humans through an increase in AROs and ARGs in human drinking supplies (**Figure 2**). Metagenomic analyses of drinking water across 25 cities on three continents show a wide diversity of ARG types in human drinking supplies with the majority of samples

having higher ARG abundance than environmental sediment and soil samples.<sup>60</sup> Altogether these studies illustrate the ubiquitous presence of AROs and ARGs in water supplies and the importance of understanding the origin and spread of ARGs in drinking water. Unless monitored and mediated, ARO and ARG environmental reservoirs can contribute to a vicious cycle that spreads AROs and ARGs over time and vast distances (**Figure 2**).<sup>54</sup>

## DISCOVERY OF NOVEL ANTIBIOTICS AND RESISTANCE MECHANISMS

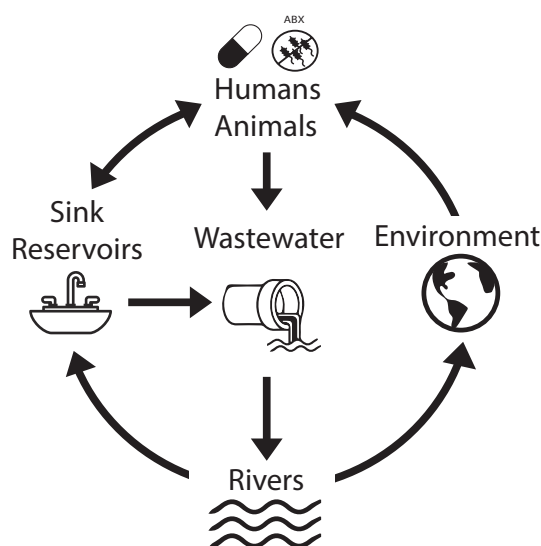
The discovery of novel antibiotic lead compounds and resistance mechanisms has accelerated with the advancements in sequencing technology.<sup>19,61–63</sup> Concurrent with these advancements, databases containing assembled genomes, annotated open reading frames, and metagenomes have expanded, which allows sequencing to serve as an emerging method for surveillance of pathogens and ARGs.<sup>64–67</sup> However, direct observation of these pathogens or ARGs is only part of the clinically relevant utility that microbial sequencing has been able to provide. This section will discuss various methods that have been used to discover new antibiotics and resistance mechanisms, or provide novel insights into previously characterized mechanisms.

## SEQUENCE-BASED METHODS

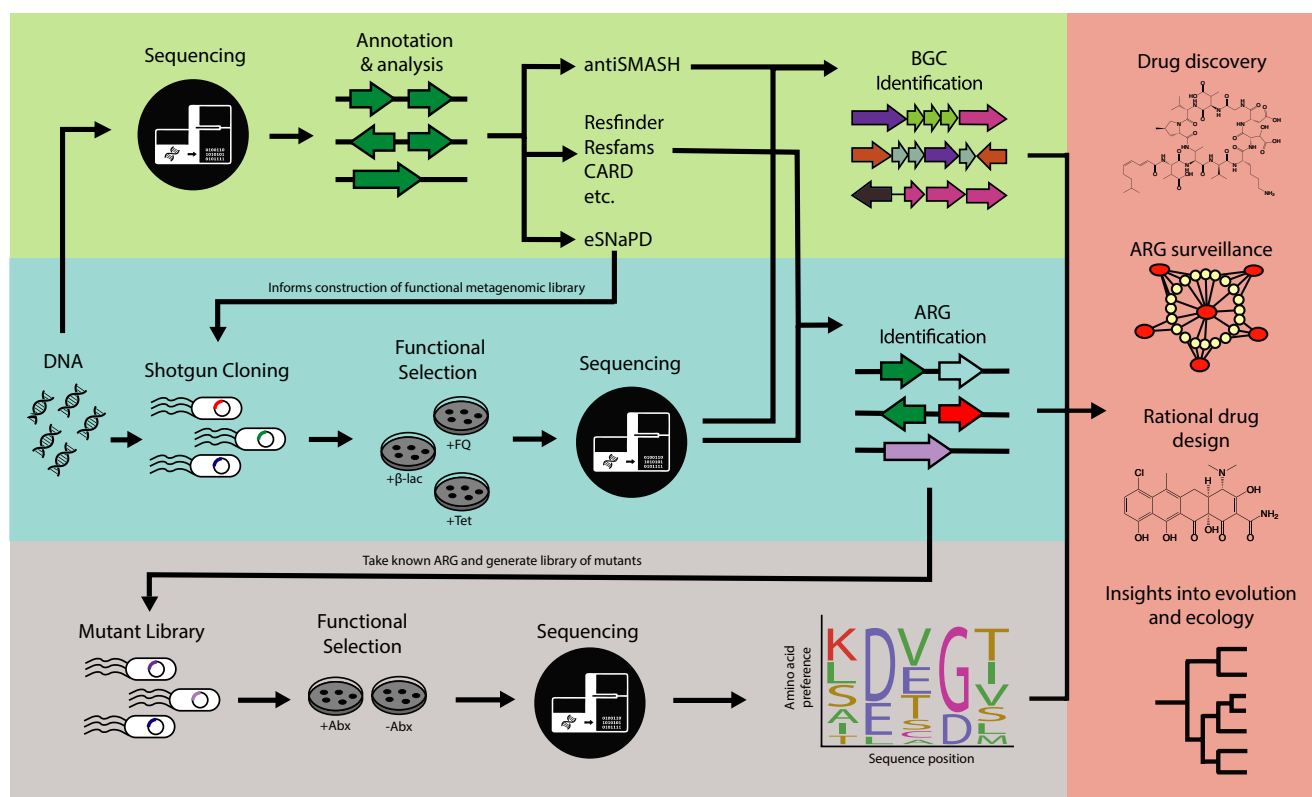
One common strategy of looking for ARGs is to use the primary nucleotide or amino acid sequence of known or putative ARGs.<sup>24,56,59,60,68</sup> These methods most commonly work by aligning a query sequence to known ARGs or markers from a curated database (**Figure 3, top**).<sup>64–67</sup> Sequence-based methods can be performed rapidly and in a massively parallel manner compared with culture-based methods of resistance determination.<sup>3</sup> Further, as new resistance mechanisms are discovered, existing data sets can be rescreened for new ARGs. However, choosing the right analysis depends on the type of sample collected and requires some expertise in bioinformatics to process and interpret the data.

## Analysis of isolates

The analysis of bacterial isolates by WGS currently proceeds primarily by highly multiplexed short-read sequencing technologies (e.g., Illumina, San Diego, CA). The reads can then be mapped to a reference genome or a database using tools like Bowtie2.<sup>69</sup> This method is advantageous because it is faster and less computationally demanding than *de novo* genome assembly (discussed below).<sup>70</sup> However, it relies on the completeness of the reference database to which the query reads are mapped. If an organism is underrepresented in the reference database, or if an ARG is novel or a distant homolog, then it will not be detected.<sup>71</sup> Further, because of the short reads (<500 bp) used in most WGS studies, some reads may spuriously align to sequences from different genes due to repetitive DNA elements and local sequence homology.<sup>72</sup> Another consequence of short read length is the loss of genomic context surrounding a particular ARG. This is of particular concern for antibiotic resistance research, because it diminishes the ability to detect transmission of ARGs through synteny with mobile genetic elements like plasmids and transposons.



**Figure 2** Antibiotic remnants from antibiotic use in humans and animals increase antibiotic concentration and antibiotic-resistant organism (ARO) and antibiotic resistance gene (ARG) abundance in wastewaters and in sink reservoirs in hospitals due to human shedding. After treatment, wastewater progresses to reservoirs such as rivers, resulting in ARG enrichment in downstream agricultural and human water sources. From rivers, ARG enrichment may be passed on to household water supplies, resulting in an ARG enrichment in sinks due to increased ARO and ARG abundance in source water. ARG enrichment in rivers also results in increased ARG abundance in sediment and other environmental water sources. This cycle can result in ARG enrichment in human pathogens and persistent environmental reservoirs of AROs and ARGs.



**Figure 3** The application of next-generation sequencing (NGS) to bacterial isolates and metagenomes has accelerated drug development by enabling the discovery of novel antibiotics and informing the rational design of drugs that reverse resistance mechanisms. Genomic data can enable intelligent surveillance methods for tracking the origin and spread of antibiotic resistance genes (ARGs), as well as insights into the evolutionary context behind emergence of a particular resistance mechanism. This can be done through sequence-based computational tools (top) or function-based methods, such as functional metagenomics methods (middle). Note the vertical arrows depicting how sequence-based and function-based methods can be combined to make inferences about a particular resistance mechanism, such as through deep mutational scanning (bottom). Abx, antibiotics; antiSMASH, antibiotics and secondary metabolite analysis shell;  $\beta$ -lac,  $\beta$ -lactam; BGC, biosynthetic gene cluster; CARD, Comprehensive Antibiotic Resistance Database; eSNAPD, environmental Surveyor of Natural Product Diversity; FQ, fluoroquinolone; tet, tetracycline.

Assembly of genomes from short reads into contiguous fragments (contigs) without the aid of a reference genome is referred to as *de novo* assembly. A recent review discusses popular genome assemblers.<sup>30</sup> The ability to analyze large contigs overcomes many of the problems of read mapping. Using assembled contigs, ARGs can be identified in the context of their surrounding genomic content, and genomes from organisms that share low sequence identity with reference sequences can be analyzed. There are limitations to this method, however. Assembly from reads is computationally demanding, requires high sequence coverage compared with read mapping, and its completeness and accuracy can be compromised by sequencing errors and repetitive DNA elements.<sup>35,70</sup> One exciting advance in this area has been the development of long-read sequencing technologies like single-molecule real-time sequencing and nanopore sequencing.<sup>73,74</sup> Instead of the 50–150 bp reads typically produced by technologies like Illumina, these methods produce reads of 10 kb to over 100 kb.<sup>73,74</sup> One advantage of these longer reads for studying antibiotic resistance of bacterial isolates is that they greatly simplify the issue of *de novo* assembly

and can even resolve a complete bacterial genome (i.e., a single contig).<sup>75</sup> Perhaps the most exciting advancement of these technologies is that they produce data much faster than short-read technologies (in as little as a few minutes compared with several days).<sup>76</sup> Nanopore sequencing in particular excels at producing data rapidly and without the need for expensive and sophisticated equipment, making it an ideal platform for point-of-care deployment. As proof of concept, Bradley *et al.*<sup>77</sup> demonstrated that antibiotic resistance profiles of *Staphylococcus aureus* could be accurately predicted using nanopore technology in as little as 7 hours.

### Analysis of metagenomes

The analysis of microbial communities by WMS is more complicated than isolate WGS analysis. Because data from WMS contains information from an unknown number of organisms with varying amounts of sequence coverage and sequencing depth, *de novo* assembly is not a trivial task.<sup>78</sup> A recent review compares popular metagenomic assemblers.<sup>79</sup> Once a genome is assembled from WMS, it can be annotated and analyzed in the same manner

as an isolate genome by WGS. However, *de novo* assembly from short-read metagenomes remains a challenge, and plasmids (which commonly harbor ARGs) are difficult to link to their host since they replicate separately from the host bacterial chromosome.<sup>80</sup> Similar to isolate genomes, long-read sequencing greatly simplifies *de novo* assembly of metagenomes. In addition to improving assemblies, long-read sequencing can better resolve plasmids from metagenomes, and simultaneous measurement of DNA methylation patterns can link plasmids to their host.<sup>80</sup> The rapid turnaround of long-read technologies has also made clinical analysis of metagenomes more feasible. Using nanopore sequencing, Břinda *et al.*<sup>76</sup> were able to accurately predict antibiotic susceptibility for *Streptococcus pneumoniae* from sputum samples within minutes, without the need for time-intensive culturing steps.

Readmapping can also be applied to metagenomic data, but this quickly becomes computationally demanding as metagenomic sequencing depth and reference database size increases. Furthermore, short readlength can lead to false positive identifications due to local sequence homology.<sup>35</sup> To overcome these problems, the ShortBRED (Short, Better Representative Extract Dataset) method was developed for assembly-free identification of ARGs in metagenomes.<sup>35</sup> ShortBRED builds short peptide sequences (markers) that are unique to a specific protein family, which are then used as a reference to map metagenomic reads. This method has been used to quickly and accurately quantify the resistome of organisms from numerous environments.<sup>62,63</sup> While most work has focused on ARGs, ShortBRED has the potential to identify and quantify any group of functionally related proteins, for example, those involved in biosynthesis of natural products.

### Search strategies

Concurrent with the proliferation of NGS technologies, the development of annotated databases of nucleotide or amino acid sequences of ARGs has allowed for the rapid screening of genomic or metagenomic data sets for resistance (**Figure 3, top**). Query sequences, whether from contigs or reads, can be aligned to sequences from known ARGs, which can be used to identify resistance mechanisms directly from sequences using pairwise sequence alignment tools like the Basic Local Alignment Search Tool (BLAST) or Bowtie2.<sup>69,81</sup>

One of the first databases for ARGs is the Antibiotic Resistance Gene Database (ARDB)<sup>64</sup> (<http://arbd.cbcb.umd.edu/>), which is a manually curated database representing >20,000 ARGs. Unfortunately, this database is not currently maintained, with the last update occurring in 2009. One actively updated database is the Comprehensive Antibiotic Resistance Database (CARD)<sup>65</sup> (<https://card.mcmaster.ca/>), which is also a manually curated database of known resistance determinants and their associated antibiotics, organized by gene ontology. CARD also offers the Resistance Gene Identifier, which is a tool with both Web-based and downloadable versions for large-scale analyses.<sup>65</sup> In addition to these general databases, several databases are maintained for specific classes of antibiotics, of note the Bush–Jacoby  $\beta$ -lactamase Lahey database and the Lactamase Engineering Database.<sup>66,67</sup> One suite of tools, ResFinder and PointFinder (<https://cge.cbs.dtu.dk/services/ResFinder/>), combines the

ARDB with numerous specialized databases to create a robust set of acquired ARGs and sequence variants known to confer resistance, which can be searched against.<sup>82</sup>

A challenge with alignment-based search methods is the risk of false negatives. Because reference databases are biased towards human pathogens and culturable bacteria, the chance of missing genes of interest when mining metagenomic data sets is very high.<sup>71</sup> One solution is to employ profile hidden Markov models (HMMs). Profile HMMs are statistical models trained on sequences with a known function that can be used to identify remote homologues and novel sequences with similar function but low pairwise identity.<sup>83</sup> Because of this ability, they have been a popular choice for discovery of novel resistance and biosynthetic genes. Resfams (<http://www.dantaslab.org/resfams>) is a curated database of HMMs that was trained using protein sequences from the CARD, Lactamase Engineering Database, and Bush–Jacoby  $\beta$ -lactamase databases.<sup>71</sup> To better capture resistance profiles from uncultured bacteria, Resfams was also trained on ARGs obtained from culture-independent functional selections from soil and human commensal microbiotas. Compared with sequence alignment, an HMM-based approach leads to significantly more sensitive and less biased screening of environmental metagenomes. When searching soil and human gut microbiomes, BLAST searches using CARD and ARDB were unable to identify 64% of the ARGs identified with Resfams.<sup>71</sup> Using a gold-standard set of annotated genes from 95 multidrug resistant soil isolates (reported in Forsberg *et al.*<sup>10</sup>), analysis with Resfams did not identify a single ARG that was not predicted by extensive manual annotation of the sequences.<sup>71</sup> This sensitivity and specificity demonstrate the ability of HMM-based approaches to accurately screen ARGs in genomes of underrepresented organisms.

Diverse environmental microbiomes and the bioactive molecules they produce represent a rich source of potential lead compounds for novel antibiotics, which may be identified using similar methods to those used for metagenomic ARG characterization. Similar to ARG identification, BLAST-based searches may not capture biosynthetic gene clusters (BGCs) present in organisms that are underrepresented in databases. To address this, the antibiotics and secondary metabolite analysis shell (antiSMASH) was developed, which integrates an HMM-based method to identify potential BGCs.<sup>84</sup> Also, antiSMASH implements many other tools, such as domain analysis, active site prediction, and chemical structure prediction. AntiSMASH is able to rapidly identify and characterize many different types of BGCs, including the important polyketide synthase and non-ribosomal peptide synthetase families. Using this pipeline, researchers were able to identify over 3,000 small-molecule BGCs in 752 metagenomes from the NIH Human Microbiome Project.<sup>85</sup> As proof of concept, the authors purified and characterized a predicted thiopeptide (an antibiotic class that was shown to be widely distributed in the microbiomes of human commensals) and demonstrated its antibacterial activity against numerous potential human pathogens, including *S. aureus*, *Enterococcus faecalis*, *Corynebacterium aurimucosum*, and *Gardnerella vaginalis*.<sup>85</sup>

Sequence-based methods are highthroughput compared with culture-based methods, and they enable analysis of organisms that



are recalcitrant to culturing. This allows for rapid and reasonably accurate screening of genomic or metagenomic data sets for functions of interest. However, large-scale analyses still require access to sophisticated sequencing equipment, massive computational resources, and expertise in bioinformatics, making their deployment to the clinic more difficult. This is a rapidly changing area, and the development of long-read sequencing technologies represents a step towards practical point-of-care deployment.<sup>76,77</sup> Furthermore, software tools are continually being improved to be accessible for a broad user base, as well as lightweight enough to run on an average laptop computer.<sup>76,77</sup> Many of the software tools discussed thus far are available free and open-source, which increases accessibility outside of large academic centers.

### FUNCTIONAL METAGENOMICS

Because sequence-based methods ultimately rely on comparisons to prior information, they are limited in their ability to identify entirely novel mechanisms. Functional metagenomics is a unique method for mining functions of interest from metagenomes in an unbiased, culture-independent way.<sup>10,86–88</sup> The basic functional metagenomic workflow involves extraction of metagenomic DNA from a community of interest (e.g., soil, human or animal gut) and cloning fragments into an expression vector, creating the functional metagenomic library. The library is then transformed into a heterologous host that can be screened for a function of interest in a high-throughput manner.<sup>87,89</sup> The gene or genes responsible for the function can be identified by sequencing the selected plasmids (**Figure 3, middle**). This method outperforms solely sequence-based methods at discovering novel genes because the selection is independent of prior knowledge about a mechanism or representation in a reference database.<sup>8,10,62,87–89</sup> More important, it allows for the identification of functions from organisms that would be otherwise extremely difficult to culture. This section will discuss cases where this method has been successfully employed to discover novel ARGs and bioactive compounds.

In the case of antibiotic resistance discovery, simply plating a functional metagenomic library on media containing antibiotics at concentrations that inhibit the selection host can select for potential novel ARGs. Several studies have used functional metagenomics to discover thousands of novel ARGs from metagenomes of diverse phylogenetic backgrounds and geographic locations.<sup>8,10,14</sup> Recently a comprehensive screen of functional metagenomic libraries was able to identify a new family of enzymes that conferred resistance to tetracycline antibiotics by chemical inactivation.<sup>89</sup> This significantly expanded the class of tetracycline-inactivating enzymes. In fact, prior to this the only biochemically characterized enzyme in this class was Tet(X), whose encoding gene (*tet(X)*) was isolated from a plasmid from the human commensal *Bacteroides fragilis*. These new enzymes (Tet(47)–Tet(55)), dubbed the tetracycline destructases, shared at most 24.4% amino acid identity to Tet(X), demonstrating the power of functional metagenomics to identify genes that share function but with low sequence identity to known genes. Comparative genomic analysis of *tet(49)*–*tet(55)* revealed another tetracycline destructase in the genome of the pathogen *Legionella longbeachae*, which was also shown to provide

resistance when expressed in a heterologous host as well as the parent pathogenic host.<sup>89,90</sup>

Identification of novel clinically relevant ARGs is ideally followed by investigations of their substrate specificity and sequence–structure–function relationships, as well as to screen for inhibitors which can enable redeployment of the targeted antibiotics. For example, x-ray crystallographic structural determination of the new soil-derived tetracycline destructases showed that while they share little primary sequence identity with Tet(X), they exhibit a similar overall three-dimensional architecture, though with a few key differences.<sup>90</sup> The authors hypothesized that the shared structural information could be used in rational design of tetracycline destructase inhibitors, analogous to the commonly used  $\beta$ -lactamase inhibitors.<sup>91</sup> This led to the identification of anhydrotetracycline as a potential lead compound for tetracycline destructase inhibition. Further, the co-crystal structure of Tet(50) with anhydrotetracycline revealed a unique binding mode where the enzyme is locked in an inactive conformation. This powerful structure–function investigation led to the conclusion that anhydrotetracycline could be coupled with tetracyclines to inhibit this class of enzymes, which was validated *in vitro* and against model pathogens expressing tetracycline destructases.<sup>90</sup> Recently these same authors described a novel synthesis scheme for generating semisynthetic anhydrotetracycline derivatives with broader inhibitor potency and enzyme specificity, including rescue of next-generation tetracyclines against soil-derived, pathogen-derived, and human gut-derived tetracycline destructases.<sup>92</sup> This functional approach to ARG discovery provided a unique, proactive opportunity to discover new therapeutic compounds to counteract an emerging threat before it explodes into the wider population, which is particularly significant given the recent identification of Tet(X) in a number of multidrug-resistant pathogens<sup>93</sup> and the even more recent clinical approvals of next-generation tetracycline antibiotics—eravacycline and omadacycline.<sup>94,95</sup>

Screening for hits in a functional metagenomic library is most convenient when the desired outcome is gain-of-function required for survival of the transformed host bacterium. Antibiotic resistance is a natural application of this method, but other studies have employed clever strategies to leverage functional metagenomics to instead discover production of novel biomolecules, including antibiotics.<sup>19</sup> One such strategy utilizes a double-agar method to screen for production of antibiotic compounds.<sup>19</sup> In this method, the heterologous host containing the library is plated on agar, where it is allowed to incubate for enough time to produce antibiotic compounds. Next, the plates are overlaid with top agar containing an indicator strain. A functional metagenomic clone that produces an antibiotic is identified by inhibition of growth of the indicator strain. This method was used to identify 12 clones from a library representing ~ 12.25 GB of DNA from a gut microbiome that were able to inhibit the growth of *E. coli*. Sequencing of the clones identified the colicin V gene cluster, which was responsible for production of the colicin V bacteriocin.<sup>19</sup>

While the studies discussed above demonstrate the power of functional metagenomics, it is not without its limitations. One major challenge is that it relies on expression of genes outside of their native hosts. Differences have been observed when expressing

the same metagenomic library in different hosts.<sup>96,97</sup> This challenge may be addressed by selecting a phylogenetically diverse group of hosts for expression of a functional metagenomic library, though this increases the cost and reduces the throughput of the system.<sup>96,97</sup> Conversely, just because a clone produces a selected phenotype of interest (e.g., antibiotic resistance) in the heterologous host, especially during forced overexpression, it does not mean that it produces the same phenotype in the original host, or indeed that it is even expressed in that native context. Accordingly, ARGs discovered from functional metagenomic selections are better categorized as illuminating functional resistome potential, which may contribute to horizontal ARG spread, rather than describing the true resistance phenotypes of the microbial members of the studied community. Culture-based phenotyping still remains the only guaranteed way to directly assess resistance of specific microbes to specific antibiotics.

### USING MASSIVELY PARALLEL SEQUENCING AND SYNTHETIC BIOLOGY AS A TOOL

Advances in NGS technology have not only made searching for interesting but uncharacterized genes much faster and easier but have opened the door to exciting new technologies that allow for better characterization of known genes or mechanisms. Methods that combine bioinformatic, microbiologic, and synthetic biology represent exciting new approaches to discover new antibiotic compounds or new insights into drug resistance.

One example of this is in the use of bioinformatics to improve access to the biosynthetic capacity of metagenomic libraries. Because BGCs often share common sequence motifs, PCR amplification of these motifs from a metagenomic sample using degenerate primers followed by amplicon sequencing allows profiling of the biosynthetic capacity of that sample.<sup>18</sup> To facilitate this, the environmental Surveyor of Natural Product Diversity (eSNaPD, available at <http://esnapd2.rockefeller.edu/>) was developed. eSNaPD is a bioinformatics platform designed for analysis of amplicons in the context of natural product biosynthesis. Using sequence tags from known BGCs, eSNaPD phylogenetically organizes amplicon sequences based on BGC content and predicted chemical output, highlighting novel or divergent biosynthetic pathways. Samples predicted to contain these novel or divergent pathways can then be prioritized for functional metagenomic library construction, expression, and screening (Figure 3, top). Recently, this method was successfully used to discover malacidins, a previously uncharacterized class of calcium-dependent antibiotics from 2,000 unique soil samples.<sup>61</sup> One extracted malacidin was shown to inhibit growth of a wide range of gram-positive bacteria, including multidrug-resistant pathogens.<sup>61</sup>

The previous example demonstrates how bioinformatics can be combined with metagenomic sequencing to explore the rich diversity of natural products that microbial communities can offer. Perhaps an even more interesting application of sequencing technologies has been to pursue diversity that is not represented in nature. The development of deep mutational scanning represents one step in that direction.<sup>98–100</sup> Deep mutational scanning is an extension of the classic experiment where a researcher

measures how specific mutations affect the function of a protein of interest, but instead of examining a handful of mutations, the researcher can simultaneously observe the effects of thousands of different protein variants simultaneously. Briefly, a mutant library is made using multiplex PCR and primers designed with random codons that amplify the target gene with every possible amino acid substitution at every possible location in the gene. Next, the library is cloned into an appropriate host, expressed, and selected for a function of interest. The cells with active protein variants are enriched and the cells with inactive variants are depleted, which is quantified by deep sequencing of the library from selected cells and unselected (control) cells.<sup>99</sup> By creating an unbiased mutant library, deep mutational scanning can reveal intrinsic protein properties, unexpected observations, and insights into the evolution of that protein (Figure 3, bottom).<sup>98</sup> Recently, deep mutational scanning was applied to the NDM-1 (New Delhi metallo-beta-lactamase 1) carbapenemase, where it was used to unveil the sequence requirements for catalysis of various  $\beta$ -Lactam antibiotics.<sup>100</sup> Interestingly, the authors found that the residues required for hydrolysis of penicillins and cephalosporins are a subset of those required for carbapenem hydrolysis. This suggests that the broad substrate profile of NDM-1 could have evolved from selective pressure exclusively from carbapenems. Further, the results imply that if an NDM-1 inhibitor is developed, it should be used in combination with carbapenems because the stricter sequence requirements for carbapenem hydrolysis reduces the number of substitutions possible that may confer inhibitor resistance.<sup>100</sup>

### CONCLUSION

The evolution and spread of antibiotic resistance are a pressing global concern. The only way to successfully combat the rise of antibiotic resistance is through a comprehensive approach that identifies, characterizes, and tracks emerging and evolving pathogens by integrating culture-based methods, genomics, and metagenomics, and by developing experimental and computational tools to identify mechanisms of resistance and develop new antibiotics to combat the evolution of antibiotic resistance.

### ACKNOWLEDGMENTS

We would like to acknowledge Alaric D'Souza, Robert F. Potter, Manish Boolchandani, and Drew Schwartz for providing feedback on this paper.

### FUNDING

This work was supported in part by awards to G.D. through the National Institute of Allergy and Infectious Diseases (<https://www.niaid.nih.gov/>), the Eunice Kennedy Shriver National Institute of Child Health & Human Development (<https://www.nichd.nih.gov/>), and the National Center for Complementary and Integrative Health (<https://nccih.nih.gov/>) of the National Institutes of Health under award numbers R01AI123394, R01HD092414, and R01AT009741, respectively.

### CONFLICT OF INTEREST

The authors declared no competing interests for this work.

### DISCLAIMER

The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

1. Aminov, R.I. A brief history of the antibiotic era: lessons learned and challenges for the future. *Front. Microbiol.* **1**, 134 (2010).
2. Brown, E.D. & Wright, G.D. Antibacterial drug discovery in the resistance era. *Nature* **529**, 336–343 (2016).
3. Adu-Oppong, B., Gasparini, A.J. & Dantas, G. Genomic and functional techniques to mine the microbiome for novel antimicrobials and antimicrobial resistance genes. *Ann. N. Y. Acad. Sci.* **1388**, 42–58 (2017).
4. Cassini, A. et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect. Dis.* **19**, 56–66 (2019).
5. Crofts, T.S., Gasparini, A.J. & Dantas, G. Next-generation approaches to understand and combat the antibiotic resistome. *Nat. Rev. Microbiol.* **15**, 422–434 (2017).
6. D'Costa, V.M. et al. Antibiotic resistance is ancient. *Nature* **477**, 457–461 (2011).
7. Bhullar, K. et al. Antibiotic resistance is prevalent in an isolated cave microbiome. *PLoS ONE* **7**, e34953 (2012).
8. Clemente, J.C. et al. The microbiome of uncontacted Amerindians. *Sci. Adv.* **1**, e1500183 (2015).
9. Knapp, C.W., Dolfing, J., Ehler, P.A. & Graham, D.W. Evidence of increasing antibiotic resistance gene abundances in archived soils since 1940. *Environ. Sci. Technol.* **44**, 580–587 (2010).
10. Forsberg, K.J., Reyes, A., Wang, B., Selleck, E.M., Sommer, M.O. & Dantas, G. The shared antibiotic resistome of soil bacteria and human pathogens. *Science* **337**, 1107–1111 (2012).
11. Poirel, L., Rodriguez-Martinez, J.M., Mammeri, H., Liard, A. & Nordmann, P. Origin of plasmid-mediated quinolone resistance determinant QnrA. *Antimicrob. Agents Chemother.* **49**, 3523–3525 (2005).
12. Wang, R. et al. The global distribution and spread of the mobilized colistin resistance gene mcr-1. *Nat. Commun.* **9**, 1179 (2018).
13. Liu, Y.Y. et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *Lancet Infect. Dis.* **16**, 161–168 (2016).
14. Forsberg, K.J. et al. Bacterial phylogeny structures soil resistomes across habitats. *Nature* **509**, 612–616 (2014).
15. Finley, R.L. et al. The scourge of antibiotic resistance: the important role of the environment. *Clin. Infect. Dis.* **57**, 704–710 (2013).
16. Douglas, A.P. et al. Utilizing genomic analyses to investigate the first outbreak of vanA vancomycin-resistant *Enterococcus* in Australia with emergence of daptomycin non-susceptibility. *J. Med. Microbiol.* **68**, 303–308 (2019).
17. Johnson, R.C. et al. Investigation of a cluster of sphingomonas koreensis infections. *N. Engl. J. Med.* **379**, 2529–2539 (2018).
18. Katz, M., Hover, B.M. & Brady, S.F. Culture-independent discovery of natural products from soil metagenomes. *J. Ind. Microbiol. Biotechnol.* **43**, 129–141 (2016).
19. Cohen, L.J., Han, S., Huang, Y.H. & Brady, S.F. Identification of the colicin V bacteriocin gene cluster by functional screening of a human microbiome metagenomic library. *ACS Infect. Dis.* **4**, 27–32 (2018).
20. Wieser, A., Schneider, L., Jung, J. & Schubert, S. MALDI-TOF MS in microbiological diagnostics-identification of microorganisms and beyond (mini review). *Appl. Microbiol. Biotechnol.* **93**, 965–974 (2012).
21. Jorgensen, J.H. & Ferraro, M.J. Antimicrobial susceptibility testing: a review of general principles and contemporary practices. *Clin. Infect. Dis.* **49**, 1749–1755 (2009).
22. Didelot, X., Bowden, R., Wilson, D.J., Peto, T.E.A. & Crook, D.W. Transforming clinical microbiology with bacterial genome sequencing. *Nat. Rev. Genet.* **13**, 601–612 (2012).
23. Köser, C.U. et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog.* **8**, e1002824 (2012).
24. Potter, R.F. et al. Population structure, antibiotic resistance, and uropathogenicity of *Klebsiella variicola*. *MBio* **9**, e02481-18 (2018).
25. Hu, Z. et al. Metagenomic next-generation sequencing as a diagnostic tool for toxoplasmic encephalitis. *Ann. Clin. Microbiol. Antimicrob.* **17**, 45 (2018).
26. Rodrigues, C., Passet, V., Rakotondrasoa, A. & Brisse, S. Identification of *Klebsiella pneumoniae*, *Klebsiella quasipneumoniae*, *Klebsiella variicola* and related phylogroups by MALDI-TOF mass spectrometry. *Front. Microbiol.* **9**, 3000 (2018).
27. Lu, Y., Feng, Y., McNally, A. & Zong, Z. Occurrence of colistin-resistant hypervirulent *Klebsiella variicola*. *J. Antimicrob. Chemother.* **73**, 3001–3004 (2018).
28. Maatallah, M. et al. *Klebsiella variicola* is a frequent cause of bloodstream infection in the stockholm area, and associated with higher mortality compared to *K. pneumoniae*. *PLoS ONE* **9**, e113539 (2014).
29. Decraene, V. et al. A large, refractory nosocomial outbreak of *Klebsiella pneumoniae* carbapenemase-producing *Escherichia coli* demonstrates carbapenemase gene outbreaks involving sink sites require novel approaches to infection control. *Antimicrob. Agents Chemother.* **62**, e01689-18 (2018). <https://doi.org/10.1128/AAC.01689-18>
30. Boolchandani, M., D'Souza, A. & Dantas, G. Sequencing-based methods and resources to study antimicrobial resistance. *Nat. Rev. Genet.* **20**, 356–370 (2019).
31. American Academy of Microbiology. *Applications of Clinical Microbial Next-Generation Sequencing: Report on an American Academy of Microbiology colloquium held in Washington, DC, in April 2015* (American Society for Microbiology, Washington, DC, 2016).
32. Quainoo, S. et al. Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. *Clin. Microbiol. Rev.* **30**, 1015–1063 (2017).
33. Truong, D.T. et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12**, 902–903 (2015).
34. Abubucker, S. et al. Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput. Biol.* **8**, e1002358 (2012).
35. Kaminski, J., Gibson, M.K., Franzosa, E.A., Segata, N., Dantas, G. & Huttenhower, C. High-specificity targeted functional profiling in microbial communities with ShortBRED. *PLoS Comput. Biol.* **11**, e1004557 (2015).
36. Guo, L.Y. et al. Detection of pediatric bacterial meningitis pathogens from cerebrospinal fluid by next-generation sequencing technology. *J. Infect.* **78**, 323–337 (2019).
37. Boucher, A. et al. Epidemiology of infectious encephalitis causes in 2016. *Med. Mal. Infect.* **47**, 221–235 (2017).
38. Shao, L., Ling, Z., Chen, D., Liu, Y., Yang, F. & Li, L. Disorganized gut microbiome contributed to liver cirrhosis progression: a meta-omics-based study. *Front. Microbiol.* **9**, 3166 (2018).
39. Opota, O., Jatón, K. & Greub, G. Microbial diagnosis of bloodstream infection: towards molecular diagnosis directly from blood. *Clin. Microbiol. Infect.* **21**, 323–331 (2015).
40. Marotz, C.A., Sanders, J.G., Zuniga, C., Zaramela, L.S., Knight, R. & Zengler, K. Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome* **6**, 42 (2018).
41. Johnson, A.P. & Woodford, N. Global spread of antibiotic resistance: the example of New Delhi metallo-beta-lactamase (NDM)-mediated carbapenem resistance. *J. Med. Microbiol.* **62**, 499–513 (2013).
42. Walsh, T.R., Weeks, J., Livermore, D.M. & Toleman, M.A. Dissemination of NDM-1 positive bacteria in the New Delhi environment and its implications for human health: an environmental point prevalence study. *Lancet Infect. Dis.* **11**, 355–362 (2011).
43. Ng, C. et al. Characterization of metagenomes in urban aquatic compartments reveals high prevalence of clinically relevant antibiotic resistance genes in wastewaters. *Front. Microbiol.* **8**, 2200 (2017).



44. Lax, S. *et al.* Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science* **345**, 1048–1052 (2014).
45. Lax, S. & Gilbert, J.A. Hospital-associated microbiota and implications for nosocomial infections. *Trends Mol. Med.* **21**, 427–432 (2015).
46. Vaz-Moreira, I., Nunes, O.C. & Manaia, C.M. Diversity and antibiotic resistance in *Pseudomonas* spp. from drinking water. *Sci. Total Environ.* **426**, 366–374 (2012).
47. Wen, X. *et al.* Occurrence and contamination profiles of antibiotic resistance genes from swine manure to receiving environments in Guangdong Province southern China. *Ecotoxicol. Environ. Saf.* **173**, 96–102 (2019).
48. Vaz-Moreira, I., Nunes, O.C. & Manaia, C.M. Bacterial diversity and antibiotic resistance in water habitats: searching the links with the human microbiome. *FEMS Microbiol. Rev.* **38**, 761–778 (2014).
49. Ruiz-Fons, F. A review of the current status of relevant zoonotic pathogens in wild swine (*Sus scrofa*) populations: changes modulating the risk of transmission to humans. *Transbound. Emerg. Dis.* **64**, 68–88 (2017).
50. Richard, M. *et al.* Factors determining human-to-human transmissibility of zoonotic pathogens via contact. *Curr. Opin. Virol.* **22**, 7–12 (2017).
51. Zhu, Y.G. *et al.* Diverse and abundant antibiotic resistance genes in Chinese swine farms. *Proc. Natl Acad. Sci. USA* **110**, 3435–3440 (2013).
52. Xie, W.Y., Yang, X.P., Li, Q., Wu, L.H., Shen, Q.R. & Zhao, F.J. Changes in antibiotic concentrations and antibiotic resistome during commercial composting of animal manures. *Environ. Pollut.* **219**, 182–190 (2016).
53. Ruan, Z. *et al.* Emergence of a ST2570 *Klebsiella pneumoniae* isolate carrying *mcr-1* and *blaCTX-M-14* recovered from a blood-stream infection in China. *Clin. Microbiol. Infect.* **25**, 916–918 (2019). <https://doi.org/10.1016/j.cmi.2019.02.005>
54. Reddy, B. & Dubey, S.K. River Ganges water as reservoir of microbes with antibiotic and metal ion resistance genes: high throughput metagenomic approach. *Environ. Pollut.* **246**, 443–451 (2019).
55. Fernandes, T., Vaz-Moreira, I. & Manaia, C.M. Neighbor urban wastewater treatment plants display distinct profiles of bacterial community and antibiotic resistance genes. *Environ. Sci. Pollut. Res. Int.* **26**, 11269–11278 (2019).
56. Guo, J., Li, J., Chen, H., Bond, P.L. & Yuan, Z. Metagenomic analysis reveals wastewater treatment plants as hotspots of antibiotic resistance genes and mobile genetic elements. *Water Res.* **123**, 468–478 (2017).
57. Czekalski, N. *et al.* Inactivation of antibiotic resistant bacteria and resistance genes by ozone: from laboratory experiments to full-scale wastewater treatment. *Environ. Sci. Technol.* **50**, 11862–11871 (2016).
58. Amos, G.C., Zhang, L., Hawkey, P.M., Gaze, W.H. & Wellington, E.M. Functional metagenomic analysis reveals rivers are a reservoir for diverse antibiotic resistance genes. *Vet. Microbiol.* **171**, 441–447 (2014).
59. Chen, H., Bai, X., Jing, L., Chen, R. & Teng, Y. Characterization of antibiotic resistance genes in the sediments of an urban river revealed by comparative metagenomics analysis. *Sci. Total Environ.* **653**, 1513–1521 (2019).
60. Ma, L. *et al.* Catalogue of antibiotic resistome and host-tracking in drinking water deciphered by a large scale survey. *Microbiome* **5**, 154 (2017).
61. Hover, B.M. *et al.* Culture-independent discovery of the malacids as calcium-dependent antibiotics with activity against multidrug-resistant gram-positive pathogens. *Nat. Microbiol.* **3**, 415–422 (2018).
62. Pehrsson, E.C. *et al.* Interconnected microbiomes and resistomes in low-income human habitats. *Nature* **533**, 212–216 (2016).
63. Gibson, M.K. *et al.* Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nat. Microbiol.* **1**, 16024 (2016).
64. Liu, B. & Pop, M. ARDB—antibiotic resistance genes database. *Nucleic Acids Res.* **37**, D443–D447 (2009).
65. Jia, B. *et al.* CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573 (2017).
66. Bush, K. & Jacoby, G.A. Updated functional classification of beta-lactamases. *Antimicrob. Agents Chemother.* **54**, 969–976 (2010).
67. Thai, Q.K., Bös, F. & Pleiss, J. The Lactamase engineering database: a critical survey of TEM sequences in public databases. *BMC Genom.* **10**, 390 (2009).
68. Zankari, E. *et al.* Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. *J. Antimicrob. Chemother.* **68**, 771–777 (2013).
69. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
70. Lin, Y., Li, J., Shen, H., Zhang, L., Papasian, C.J. & Deng, H.W. Comparative studies of *de novo* assembly tools for next-generation sequencing technologies. *Bioinformatics* **27**, 2031–2037 (2011).
71. Gibson, M.K., Forsberg, K.J. & Dantas, G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* **9**, 207–216 (2015).
72. Bocanegra-Ibarias, P. *et al.* Molecular and microbiological report of a hospital outbreak of NDM-1-carrying Enterobacteriaceae in Mexico. *PLoS ONE* **12**, e0179651 (2017).
73. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009).
74. Clarke, J., Wu, H.C., Jayasinghe, L., Patel, A., Reid, S. & Bayley, H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* **4**, 265–270 (2009).
75. Liao, Y.C., Lin, S.H. & Lin, H.H. Completing bacterial genome assemblies: strategy and performance comparisons. *Sci. Rep.* **5**, 8747 (2015).
76. Břinda, K. *et al.* Lineage calling can identify antibiotic resistant clones within minutes. *bioRxiv* (2018). <https://doi.org/10.1101/403204>
77. Bradley, P. *et al.* Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nat. Commun.* **6**, 10063 (2015).
78. Ghurye, J.S., Cepeda-Espinoza, V. & Pop, M. Metagenomic assembly: overview, challenges and applications. *Yale J. Biol. Med.* **89**, 353–362 (2016).
79. van der Walt, A.J., van Goethem, M.W., Ramond, J.B., Makhalanyane, T.P., Reva, O. & Cowan, D.A. Assembling metagenomes, one community at a time. *BMC Genom.* **18**, 521 (2017).
80. Beaulaurier, J. *et al.* Metagenomic binning and association of plasmids with bacterial host genomes using DNA methylation. *Nat. Biotechnol.* **36**, 61–69 (2018).
81. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
82. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **67**, 2640–2644 (2012).
83. Eddy, S.R. Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).
84. Blin, K. *et al.* antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45**, W36–W41 (2017).
85. Donia, M.S. *et al.* A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158**, 1402–1414 (2014).
86. Rondon, M.R. *et al.* Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* **66**, 2541–2547 (2000).
87. Pehrsson, E.C., Forsberg, K.J., Gibson, M.K., Ahmadi, S. & Dantas, G. Novel resistance functions uncovered using functional metagenomic investigations of resistance reservoirs. *Front. Microbiol.* **4**, 145 (2013).
88. Moore, A.M. *et al.* Pediatric fecal microbiota harbor diverse and novel antibiotic resistance genes. *PLoS ONE* **8**, e78822 (2013).



89. Forsberg, K.J., Patel, S., Wencewicz, T.A. & Dantas, G. The tetracycline destructases: a novel family of tetracycline-inactivating enzymes. *Chem. Biol.* **22**, 888–897 (2015).
90. Park, J. et al. Plasticity, dynamics, and inhibition of emerging tetracycline resistance enzymes. *Nat. Chem. Biol.* **13**, 730–736 (2017).
91. Drawz, S.M. & Bonomo, R.A. Three decades of beta-lactamase inhibitors. *Clin. Microbiol. Rev.* **23**, 160–201 (2010).
92. Markley, J.L. et al. Semisynthetic analogues of anhydrotetracycline as inhibitors of tetracycline destructase enzymes. *ACS Infect. Dis.* **5**, 618–633 (2019).
93. Leski, T.A. et al. Multidrug-resistant tet(X)-containing hospital isolates in Sierra Leone. *Int. J. Antimicrob. Agents* **42**, 83–86 (2013).
94. Zhanel, G.G. et al. Review of eravacycline, a novel fluorocycline antibacterial agent. *Drugs* **76**, 567–588 (2016).
95. Tanaka, S.K., Steenbergen, J. & Villano, S. Discovery, pharmacology, and clinical profile of omadacycline, a novel aminomethylcycline antibiotic. *Bioorg. Med. Chem.* **24**, 6409–6419 (2016).
96. Martinez, A. et al. Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Appl. Environ. Microbiol.* **70**, 2452–2463 (2004).
97. Craig, J.W., Chang, F.Y., Kim, J.H., Obiajulu, S.C. & Brady, S.F. Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Appl. Environ. Microbiol.* **76**, 1633–1641 (2010).
98. Fowler, D.M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **11**, 801–807 (2014).
99. Fowler, D.M., Stephany, J.J. & Fields, S. Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nat. Protoc.* **9**, 2267–2284 (2014).
100. Sun, Z., Hu, L., Sankaran, B., Prasad, B.V.V. & Palzkill, T. Differential active site requirements for NDM-1 beta-lactamase hydrolysis of carbapenem versus penicillin and cephalosporin antibiotics. *Nat. Commun.* **9**, 4524 (2018).