REVIEW

Approaches for characterizing and tracking hospital-associated multidrug-resistant bacteria

Kevin S. Blake[1] · JooHee Choi[1] · Gautam Dantas[1,2,3,4]

Abstract

Hospital-associated infections are a major concern for global public health. Infections with antibiotic-resistant pathogens can cause empiric treatment failure, and for infections with multidrug-resistant bacteria which can overcome antibiotics of "last resort" there exists no alternative treatments. Despite extensive sanitization protocols, the hospital environment is a potent reservoir and vector of antibiotic-resistant organisms. Pathogens can persist on hospital surfaces and plumbing for months to years, acquire new antibiotic resistance genes by horizontal gene transfer, and initiate outbreaks of hospital-associated infections by spreading to patients via healthcare workers and visitors. Advancements in next-generation sequencing of bacterial genomes and metagenomes have expanded our ability to (1) identify species and track distinct strains, (2) comprehensively profile antibiotic resistance genes, and (3) resolve the mobile elements that facilitate intra- and intercellular gene transfer. This information can, in turn, be used to characterize the population dynamics of hospital-associated microbiota, track outbreaks to their environmental reservoirs, and inform future interventions. This review provides a detailed overview of the approaches and bioinformatic tools available to study isolates and metagenomes of hospital-associated bacteria, and their multi-layered networks of transmission.

Keywords Antibiotic resistance · Next-generation sequencing · Microbiome · Resistome · Mobilome · Hospital ICU

Introduction

Over the course of several days, the owner of the quintessentially gloomy Gothic mansion, the House of Usher [1], quietly confesses to the Narrator his belief that the source of the mysterious illness which has plagued him and his

Kevin S. Blake and JooHee Choi contributed equally.

✉ Kevin S. Blake
   kevin.blake@wustl.edu

✉ Gautam Dantas
   dantas@wustl.edu

1   The Edison Family Center for Genome Sciences & Systems
    Biology, Washington University School of Medicine,
    St. Louis, MO 63110, USA

2   Department of Molecular Microbiology, Washington
    University School of Medicine, St. Louis, MO 63110, USA

3   Department of Pathology and Immunology, Washington
    University School of Medicine, St. Louis, MO 63110, USA

4   Department of Biomedical Engineering, Washington
    University in St. Louis, St. Louis, MO 63130, USA

family for generations—is the house itself. He is possessed by the idea that his house is alive and has exerted its "silent, yet importunate and terrible influence" by repeatedly infecting his family with an unexplainable disease. While such supernatural forces are securely confined to the pages of Poe's short stories, it is possible for a building to repeatedly infect people with the same pathogens. Fortunately, this is a problem which can be addressed far less dramatically than dumping the building into a lake.

Nosocomial, or hospital-acquired infections (HAI, also referred to more generally as "healthcare-associated infections") are caused by pathogens whose source is within the hospital [2]. These pathogens can contaminate and persist on high-contact surfaces such as light switches, call buttons, and bedside rails, as well as the building's plumbing, and they can be transmitted between patients via healthcare workers and visitors [3] (Fig. 1a). In United States (US) acute care hospitals, 4% of patients have one or more HAIs, with 11.5% of these patients dying during hospitalization [4]. Critically-ill and immunocompromised patients in intensive care units (ICU) are the most vulnerable to HAIs due to severe underlying diseases and indwelling devices that
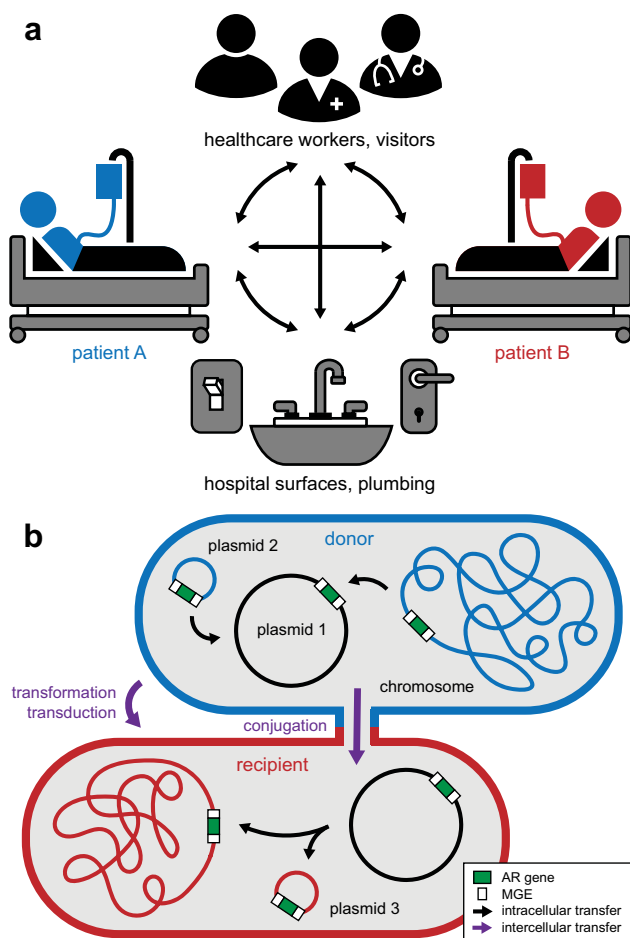
Fig. 1 Multi-layered transmission networks of nosocomial pathogens and AR genes. During outbreaks of HAIs, **a** MDR pathogens can be spread between patients via healthcare workers and visitors, and/or persistent reservoirs of hospital-associated microbiota such as high-touch surfaces and plumbing. Additionally, **b** AR genes can be spread intracellularly (e.g., chromosome to plasmid, or plasmid to plasmid) or intercellularly between bacteria of different taxa via plasmids and other MGEs.

provide an entryway for pathogens [5–7]. A study of 15,202 patients worldwide reported that 21% of ICU patients had an ICU-acquired infection, and ICU-acquired infections were independently associated with higher risk of in-hospital mortality compared with community-acquired infection [8].

Of special concern are HAIs caused by antibiotic-resistant (AR) organisms [9, 10]. The widespread use of antibiotics and antimicrobials in hospitals, with 50% of hospitalized patients and 70% of ICU patients receiving at least one antibiotic during their stay [4, 8], make these buildings a hotspot for the evolution of antibiotic and multidrug-resistant (MDR) pathogens [11]. In the US, more than 2.8 million AR infections occur each year, resulting in more than 35,000 deaths [12]. Between 2007 and 2015, the European Union saw a 2.5-fold increase in the number of infections

and deaths attributed to AR organisms [13]. Central to the efforts to combat AR is the recognition that it is a product of selective pressures present in nearly all ecological habitats, driven by diverse bacterial taxa via multiple molecular mechanisms [14].

Studies of hospital-associated microbiota are guided by three main questions: (1) What species are present? (2) What AR genes do they have? and (3) Can those genes be mobilized? These questions can be targeted to isolates of known HAI-causing pathogens, or comprehensively surveyed using patient and hospital environment metagenomic samples. Further, longitudinal sampling coupled with comparative genomics can track transmission of MDR bacteria (Fig. 1a), as well as the mobilization of their AR features between species and genetic elements (Fig. 1b), allowing researchers to track the multiple layers of strain and gene transmission during HAI outbreaks. Recent developments in next-generation DNA sequencing (NGS) technologies as well as question-specific approaches have empowered researchers to analyze genomic and metagenomic datasets at the high-resolution needed to accurately track the transmission of MDR bacteria and their genetic elements. This work can, in turn, inform infection prevention practices and gauge future outbreak risks [3].

In this review, we will summarize the approaches and bioinformatic tools available to study hospital-associated MDR bacteria, evaluate their strengths and weaknesses, and provide examples of their successful use.

Surveying taxonomy and tracking strains

Global public health organizations prioritize HAI-causing and MDR bacteria in different ways, but they all include the same key players. The Infectious Disease Society of America terms the pathogens responsible for the majority of HAIs, ESKAPE pathogens [15, 16]. ESKAPE stands for *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species. It also refers to these species' abilities to "escape" killing by antibiotics [17]. (Some groups also include *Escherichia coli*, making the acronym ESKAPEE). The US Centers for Disease Control (CDC) prominently features ESKAPE pathogens in its list of 18 AR pathogens [12], categorized based on level of concern to human health—urgent, serious, or concerning—as does the World Health Organization's (WHO) priority list of 12 pathogens [18], categorized based on the need for new antibiotics to treat them—critical, high, or medium. Given their propensity for causing HAIs, hospital surveillance efforts should pay special attention to ESKAPE pathogens.

However, ESKAPE and other nosocomial pathogens represent only a small fraction of the microbial diversity in

the hospital environment (relative abundance < 0.5%) [19]. While regular cleaning and disinfecting protocols keep the microbiota in the built environment of the hospital much less taxonomically diverse than other environments, hospitals are far from sterile [20, 21]. Metagenomics have quickly altered the field of microbiology by shifting focus from bacteria as isolated players, to members of a dynamic community which actively exchanges genetic materials and competes against one another. Here, we refer to the hospital microbiome as the microorganisms inhabiting the hospital environment (e.g., surfaces, plumbing). This includes non-pathogenic "environmental" bacteria which may not be the root cause of clinical infections, but can influence the pathogens that do. The hospital microbiome is a reservoir of AR genes that can be exchanged between environmental and disease-causing pathogens via conjugative plasmids [20], and can facilitate the persistence of MDR bacteria through the formation of biofilms [22]. Another potent reservoir of AR genes and hotspot for genetic exchange is patient-associated microbiomes (e.g., gut and skin microbiomes). These can seed the hospital environment with MDR bacteria, which can then be transmitted to other patients [21], making the hospital environment is both a reservoir and vector of AR genes and HAI-causing pathogens (Fig. 1a). Therefore, any survey of the taxonomic and functional diversity of hospital-associated microbiota is woefully incomplete if it does not consider the complete hospital and patient microbiomes.

Historically, technological limitations forced researchers to choose between one group or the other—a detailed picture of key pathogens without consideration for the "unculturable majority," or a complete but low-resolution sketch of the patient or hospital microbiome. However, by applying multiple approaches and technologies in parallel, recent studies have been able to construct the most complete picture of hospital-associated microbiota, paying appropriate attention to both the pathogenic and environmental bacteria. Below, we describe each of these technologies, the kinds of questions they can be applied to, and their limitations.

Identification of single taxon isolates

The most basic method of identifying a bacterial species is through selective culture. Here, a clinical or environmental sample is streaked onto agar-based media containing specific nutrients and/or antibiotics, and then incubated aerobically or anaerobically. The combination of these factors supports the growth of specific species, with a heavy bias for ESKAPE and other human pathogens. Selective culturing alone can be used to identify species, or as a preliminary step before more in-depth characterization. However, selective culture assays are fundamentally limited by the growth rate of the bacteria, thereby forcing microbiologists, clinicians, and patients to wait for the results, and subsequent

modifications to treatment. This delay can be extended by days or weeks if further testing (e.g., antibiotic susceptibility or biochemical testing) is needed. Additionally, species that are slow-growing, at low abundance, or simply cannot grow in the conditions tested will be overlooked by this method. Alternatively, for fastidious microorganisms whose culture is either impractical or unreliable (e.g., pathogens associated with sexually transmitted diseases), nucleic acid amplification testing (NAAT) targets specific genetic sequences for amplification by polymerase chain reaction (PCR) and fluorescent probe-based detection [23, 24].

Within the past decade, clinical microbiology laboratories have begun a revolutionary shift away from solely relying on selective culture-based identification to adopting matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) [25]. MALDI-TOF takes advantage of the fact that each species has a unique assortment and composition of cell products. This results in a distinct mass-to-charge (m/z) ratio pattern, which can then be compared to a large database of characterized organisms to match the sample with a known species. This approach enables the rapid identification of genera and species from isolates or even direct clinical samples. For example, *Mycobacterium* are notoriously slow and fastidious growers, typically taking 7–21 days to grow by conventional culture-based identification methods, but MALDI-TOF has reduced the time to identification from weeks to hours [26]. However, the principal crux of MALDI-TOF is its reliance on a pre-defined database. This is not a problem for well-studied species, such as HAI pathogens, but it can lead to the misidentification of rarer or less characterized species. For example, in one study MALDI-TOF misidentified 27/289 isolates, including isolates belonging to a novel genus of MDR *Enterobacteriaceae*, *Superficieibacter electus* [21]. To overcome such limitations, the databases for MALDI-TOF MS are constantly undergoing updates and expansions for improved characterization of diverse taxa and environments [27–29].

While the above methods can identify taxa using growth phenotypes, unique DNA fragments, or biophysical properties, the most high-resolution picture of a microorganism comes from sequencing its genome. The ever-decreasing costs of NGS technologies [30] have enabled their adoption by microbiology labs for isolate identification. Whole genome sequencing (WGS) enables the de novo assembly of bacterial isolate genomes, and study of their genetic contents and architectures [31, 32]. Briefly, DNA is extracted from the isolate sample, fragmented into the appropriate length for the sequencing platform, tagged with unique barcodes, and sequenced. Each sequenced fragment constitutes a read, and overlapping reads can be assembled into longer stretches of contiguous DNA sequences, called contigs. Assembly of this WGS data into genomes is accomplished

using bioinformatic tools such as Velvet [33], SPAdes [34], and Unicycler [35]. Other tools, such as Quast [36], can then assess the quality of an assembly by looking at several factors including the length distribution of contigs, percentage of unresolved bases, and overall coverage of the genome. Species assignments can then be made in silico using MASH [37], or, for more in-depth characterization (e.g., if MASH doesn't agree with the MALDI-TOF assignment), RNAmmer to identify the 16S sequence [38] (described in detail in the next section) followed by submission to the EzBioCloud taxonomic database [39]. The genome's open reading frames can then be identified and annotated using Prokka [40].

WGS has been used to track multiple outbreaks of MDR organisms, including vancomycin-resistant *E. faecium* (VREfm) [41], MDR *Sphingomonas koreensis* [42], and *K. pneumoniae* carbapenemase (KPC)-producing *E. coli* [43]. Additionally, WGS can be a powerful tool to confirm or re-evaluate species identifications of clinical samples made by more traditional methods. For example, *K. variicola*, a relative of the more well-known ESKAPE pathogen *K. pneumoniae*, has quickly become recognized as an emerging pathogen in its own right, and *K. variicola*-infected patients have a higher mortality than *K. pneumoniae*-infected [44]. However, WGS analyses revealed that 2–10% of isolates designated as *K. pneumoniae* by selective culture and MALDI-TOF had been misidentified, and were actually *K. variicola* [45–47].

Studies of isolate genomes are more accurate—and therefore more useful for strain tracking (see later section)—if the reads can be assembled into a complete, contiguous sequence. However, the process of fully assembling a genome using short-read WGS data alone is notoriously difficult because of highly repetitive elements that exist at multiple sites throughout the chromosome. These challenges can be mitigated through the use of long-read sequencing technologies. Platforms such as single-molecule real-time sequencing (SMRT or PacBio) and nanopore sequencing, can produce reads of > 10 kb to over 100 kb [48, 49]. As a result, they generate fewer, longer contigs per genome, which span the repeat elements that confound WGS assembly. As of the time of writing, long-read sequencing of a bacterial genome is about twice as expensive as short-read. While still affordable for a small subset of genomes, long read sequencing can become prohibitively expensive for large-scale projects involving hundreds of isolates. For this reason, most researchers choose to perform long-read sequencing on subsets of isolates—a decision typically guided by first performing WGS or antibiotic susceptibility test (AST) (described in detail in a later section) to identify isolates of interest. Long-read technologies are characteristically less accurate (75–90%) than short-read Illumina sequencing (99.8%) [48, 50]. While the errors of PacBio sequencing are random and therefore can be corrected by increasing the coverage, the errors of nanopore sequencing are non-random (i.e. occur in specific nucleotide patterns) [51]. Nevertheless this can be mitigated with hybrid assembly which uses both short- and long-reads, or where long-read assembly is followed by short-read error correction [52] using tools like Unicycler [35] and hybridSPADES [53], respectively. Even with low coverage, long-read hybrid assembly is able to resolve repetitive elements and gaps in the relatively inexpensive short reads, resulting in a cost-effective mode of obtaining high-quality, closed genomes [54].

Metagenomic approaches for community surveys

To survey the entire taxonomical landscape of the hospital microbiome including the unculturable majority, many studies have turned to metagenomic techniques. The 16S rRNA gene is highly conserved within bacterial species [55], making it an ideal candidate for phylogenetic studies [56]. In 16S rRNA gene amplicon (16S) sequencing [57, 58], conserved regions of this gene are amplified through PCR and then sequenced. Sequenced reads are assigned to bacterial taxa by comparing reads against known databases [59], or by grouping similar sequences into Operational Taxonomic Units (OTU) [60]. However, using 16S data to estimate the relative abundances of taxa can be potentially misleading, as DNA extraction and PCR amplification steps are prone to bias [61]. This can be mitigated through the use of amplicon sequence variants (ASV), which can denoise sequencing data of the amplification and sequencing errors, and then be organized into OTUs or be used as their own unit of analysis [62]. Moreover, by only sequencing a small fragment of the bacterial genome (i.e. the rRNA gene), 16S sequencing is limited to species-level resolution and is blind to other gene functions of interest, such as AR. Some of the earliest efforts to characterize hospital-associated microbiota using high-throughput metagenomic sequencing involved the use of 16S rRNA in neonatal intensive care units (NICU) [63]. For example, a study examining the guts of premature infants showed that the dominating taxa were similar to those on hospital surfaces, especially feeding and intubation tubing [64].

While WGS is limited to single-genome isolates, and 16S sequencing only sequences small fragments of the metagenome, whole metagenome shotgun (WMS) sequencing conducts short-read sequencing on all DNA present in a sample. This facilitates the characterization of both the taxonomic and functional makeup of complex metagenomes. The resolution of a WMS dataset depends on the number of sequenced reads and the complexity of the sample (i.e. the higher the resolution and/or the greater the complexity, the more reads that are required). However, the more reads generated per sample, the more expensive the sequencing becomes. Deep-whole metagenome shotgun sequencing can

resolve strain-level taxonomic and functional information, but the number of reads required (e.g., 2.5 billion reads for "ultradeep" sequencing) may be prohibitively expensive for large-scale studies [65]. On the other hand, "shallow" shotgun sequencing uses as few as 0.5 million reads, but can still resolve species-level taxonomic information and functional profiling [65]. The cost of shallow shotgun sequencing is comparable to 16S, but provides more information, at a higher resolution, with better reliability. After sequencing, taxonomic profiling of the metagenomic sample (such as composition and relative abundances) can then be determined using tools such as MetaPhlAn2 [66] or Kraken [67], and functional profiling of the metagenome at the species level with HUMAnN2 [68].

WGS and WMS sequencing are not mutually exclusive and can be used in concert. For example, WGS and WMS have been used together to identify the gut microbiota as the source of several blood-stream infections in immunocompromised hematopoietic cell transplantation recipients [69]. Here, the genomes of infection-causing blood isolates of *E. coli* and *K. pneumoniae* were matched to closely-related strains in the patients' stool—in one instance, with zero single nucleotide polymorphisms (SNPs) between the bloodstream isolate and stool strain—strongly suggesting the gut was the source of this pathogen.

Taking a step further, the de novo assembly of metagenome-assembled genomes (MAG) has been a longstanding goal in microbiome research. Assembly using WMS data is more complicated than for single-isolate WGS because the algorithms need to account for unknown abundances of different organisms with unknown phylogenetic relationships [70]. To overcome these challenges, several metagenome-specific assemblers have been developed for use with short-read sequencing data, including MEGAHIT [71], MetaSPAdes [72] and MetaVelvet [73, 74]. However, the limitations of short reads coupled with microbiome challenges—low species abundance, high strain diversity, and low recovery rates for some phyla—means this approach usually generates incomplete draft genomes of varying quality [75] by binning similar contigs based on sequence composition and coverage [76, 77]. However, as binning quality relies on the size and contiguity of the assembly (i.e. fewer longer contigs per genome increases sensitivity and specificity), this effort can be improved with long-read sequencing. For example, Lathe is a workflow for extracting high molecular weight DNA from stool, followed by long-read assembly with short-read error correction [78]. Alternatively, hybrid metagenomic assemblers which combine long reads with short reads, such as OPERA-MS [79], have been shown to provide strain-resolved genomes with greater base pair accuracy and contiguity than non-hybrid assemblies, and may be preferred where low-coverage long-read sequencing data is available.

Despite the advantages of metagenomic sequencing over the laborious process of large-scale isolation of culturable pathogens, it often cannot provide detailed genomic characterization of the key pathogens that drive HAIs because these are generally present at low relative abundances (< 0.5% on hospital surfaces) [19]. The middle-ground between these approaches is culture-based enrichment (aka quasi-metagenomics, sweep metagenomics), whereby a metagenomic sample is first enriched for target organisms or phenotypes using growth medium, and sweeps from agar plates are sequenced—as opposed to single colonies. To shift the distribution of a metagenomic sample away from abundant environmental or commensal species and toward low-abundance drug-resistant pathogens, samples can first be enriched in broth and then the mixed culture be plated on antibiotic-containing media [19]. Alternatively, to infer the population structure of a target organism, samples can be plated directly onto antibiotic-free agar [80]. Culture-based enrichment and nanopore sequencing has been used on hospital surface samples, generating 2,347 high-contiguity genomes, 1,693 phage sequences, and 5,910 closed plasmid sequences (1,400 containing AR genes), > 60% of which were novel [19]. This led to the observation that, while MDR organisms were widely distributed and persistent across sites, the bacterial communities associated with high-touch surfaces (e.g., bed rail, bedside locker, cardiac table) are taxonomically distinct from hospital plumbing (e.g., sink traps and aerators), which are known reservoirs of outbreak-associated pathogens [20, 81, 82]; however, many specific pathogens and AR genes are more common in the high-touch sites. Further, the two community types harbor distinct AR genes, with the high-touch surfaces carrying a wider diversity at lower frequencies—underscoring the distinctness of hospital environments as AR gene reservoirs.

Tracking strains across samples

Studies of HAI outbreaks are not only motivated by questions about the identity or composition of individual samples, but also by how they compare to samples from other sites or patients, and change over time. Once the identity or composition of a sample is determined using the above methods, comparative genomics can then be used to evaluate their similarity to other samples. Further, as outbreaks can be caused either by the repeated transmission of a single strain or the sporadic introduction of several distinct strains, determining the level of relatedness between genomes sampled from different sites and patients at different timepoints can provide insights into the epidemiology of an outbreak and inform decisions about how to best intervene.

Evaluating genome similarity at the species level typically involves pairwise comparisons and quantification of average nucleotide identity (ANI), the average percentage

of nucleotides which are identical between two genomes. An ANI of least 96% is considered the minimum threshold for two samples to be considered the same species [83], and ever more stringent thresholds have been used to define strains (> 99.9% ANI) [42], strain derivatives (> 99.99% ANI) [20], and direct strain transfers (> 99.999%) [84]. However, at these sub-species levels, the resolution of ANI diminishes and thus may no longer be appropriate. While there is no defined consensus, most studies rely on some form of genomic distance cutoff—a maximum threshold for mutational divergence between two samples. (Table 1 is a list of studies and their selected methods for identifying and tracing strains.)

A fundamental question for determining mutational distance is whether to compare the whole genome or just the core genes. The core genome refers to the subset of genes that are shared among most or all strains of a species [85]. Because they are evolutionarily conserved, mutations in core genes often reflect true phylogenetic relationships [86]. Tools such as Roary [87] allow this core gene/accessory gene distinction, which can then be applied using other tools for downstream analyses such as RAxML [88] to construct phylogenetic trees, and fastGEAR [89] and BAPS [90] to distinguish strains. However, core genome distance alone is somewhat limited in that it willfully ignores intergenic regions and any gene that is not present in all samples, potentially losing information about accessory genes and intergenic regions which can be important for phylogeny. Nevertheless, core- and whole-genome alignments are significantly correlated [21], making core genome alignment a computationally efficient mode of comparing conserved functions.

Once a method for genome alignment has been chosen, relatedness is typically determined by SNP distances between the isolates. In practice, strains are generally understood as sub-groups within a species that share genotypic and phenotypic characteristics [91]. However, there is no universal SNP distance threshold that defines strain-level relatedness, and as such the cutoff varies from study to study [92]. Stringent cutoffs, such as less than 10 SNPs, allow for researchers to argue for clonality [69]. However, other studies have set their cutoffs at 20 to 100 SNPs [93–95] (Table 1). In general, the literature has been permissive towards independently determined cutoff values so long as they are reasonably conservative about claiming clonality, and establish a clear numerical gap between those that are related and those that are not. A good way to do this is to first calculate all pairwise distances between isolates [21], and identify a natural "elbow" in the curve. This will be the point at which the distribution splits into two, and most likely (in a multi-patient study) where within-patient and between-patient comparisons separate. For example, in a comparative study of the gut commensal *Bacteroides fragilis*, isolates from the same patient differed by fewer than 100 SNPs, while those from different patients were on average more than 10,000 SNPs different [96]. Metadata about the isolates' sampling site or timepoint can also provide additional context and biological support for interpreting distance cutoffs.

Another factor to consider when comparing genomes what will be used as the reference genome. Many studies choose to use a well-annotated, external reference that is known to be closely related to the samples at hand [97]. For instance, in a study on chronic infections of *Burkholderia*

Table 1 Case examples of distance-based approaches to identify strains

Study	Organism	Sequencing approach	Reference	Strain cutoff
Silva et al. [95]	*Burkholderia multivorans*	WGS	Temporally initial isolate draft genome	~20 SNPs
Snitkin et al. [232]	KPC-producing *Klebsiella pneumoniae*	WGS	Temporally initial isolate	40 SNPs
Tamburinin et al. [69]	*Escherichia coli* and others	WGS and WMS	Best quality isolate draft genome	1 SNP
Coll et al. [94]	Methicillin-resistant *Staphylococcus aureus* (MRSA)	WGS	Core genome	50 SNPs
Smibert et al. [233]	Methicillin-resistant *Staphylococcus aureus* (MRSA)	WGS	Core genome with publicly available reference genome	183 SNPs
Guerra-Assunção et al. [234]	*Mycobacterium tuberculosis*	WGS	Publicly available reference genome	10 SNPs
Clark et al. [93]	Multidrug-resistant *Mycobacterium tuberculosis*	WGS	Publicly available reference genome	50 SNPs
Donkor et al. [235]	*Staphylococcus aureus*	WGS	Core genome with publicly available reference genome	2 SNPs
Sharma et al. [236]	*Streptococcus pyogenes*	WGS	Publicly available reference genome	1 SNP
Sundermann et al. [237]	Vancomycin resistant *Enterococcus* (VRE)	WGS	Publicly available reference genome	15 SNPs

multivorans in cystic fibrosis patients, the authors annotate polymorphisms within a given population by referring to the well-annotated AU0158 reference genome derived from the same outbreak [97]. However, this approach relies on the assumption that their isolates are closely related to the reference and may lose resolution on genomic regions unique to the samples. In contrast, studies with longitudinally collected samples may choose to use the first isolate obtained as the reference, and interpret subsequent isolates as representing mutations accumulated since the initial infection event [95]. This longitudinal interpretation should be approached with caution, however, as it relies on sampling methods that are representative of the diversity within the cohort, as well as accurate representation of the initial infection event itself. What may appear to be a longitudinal evolutionary trajectory may actually be the coexistence of multiple strains within a sample set that are being sporadically sampled and sequenced. Still another potential reference is simply to select the best quality (i.e., the most contiguous and complete) draft genome [69]. These approaches are not mutually exclusive, and some studies have opted to take a multi-step approach: first using a publicly available genome as reference, building a phylogeny demonstrating that clades are grouped by patient, and then constructing patient-specific draft genomes for higher resolution analyses [96]. As such, it is advised that researchers select a reference genome appropriate for the biological questions at hand, while maintaining a high assembly quality and rigorous sampling practices.

Accurate quantification of genomic distances in these ways requires high-quality assemblies, but other methods for resolving strains can be used when this is not available. Multilocus sequence typing (MLST) is a simple, curated way to consider strain types without introducing an arbitrary mutational threshold. MLSTs are defined on a per-species basis as specific mutational patterns on a set of five to seven housekeeping genes. Strain types (STs) have been defined for more than 100 microbial species and genera, making it the gold standard of typing [98, 99]. While traditionally determined through targeted PCR and Sanger sequencing, recent advancements in NGS have enabled the accurate in silico typing of samples from WGS [100]. Alternatively, some species, such as *E. coli,* are more broadly categorized and referred to in terms of phylogenetic groups (phylogroups). Phylogroups are members of the same species, but differ in biological characteristics such as ecological niche and virulence [101, 102]. Like MLST, phylogroups had been originally determined by PCR [103], but in silico predictions of phylogroup based on WGS are also available [104].

Surveying hospital-associated microbiota is a complex endeavor which can be addressed using several approaches. The method(s) chosen will ultimately depend on the available resources (e.g., sample types, analytical expertise, access to equipment) as well as the biological questions at-hand. For example, a study interested in the long-term adaptations of a pathogen within a host or environmental site would benefit from longitudinal sampling of multiple isolates per timepoint with WGS, while a similar study expecting to observe cross-species and cross-genera interactions may instead opt for WMS sequencing. Additionally, a study interested in surveilling pathogens in the hospital environment could sample many patients and surfaces, but limit the total number of samples per person and surface or focus exclusively on MDR organisms. Regardless, a great strength to NGS-based approaches is that once sequenced, the resulting dataset can be shared publicly and re-analyzed using different bioinformatic tools to answer other questions (e.g., AR gene repertoire, presence of MGEs).

Profiling antibiotic resistance

Effective surveillance of hospital-associated microbiota requires knowledge of not only which species are present, but also what their antibiotic resistances and susceptibilities are. AR genes have existed in bacterial populations for tens of thousands of years [105], but industrial production and extensive anthropogenic use of antibiotics has imposed immense selection pressures which results in the amplification, diversification, and dissemination of resistance on a global scale [106]. AR is usually genetically encoded and can arise through mutations such as gene duplication [107], overexpression and point mutations [108], or the acquisition of fully-functioning genes via horizontal transfer [108, 109]. (Non-genetic mechanisms of resistance, such as tolerance and persistence, have been reviewed in [110, 111].) Mechanistically, antibiotic resistance can be achieved in three ways: (1) keeping the antibiotic out of the cell through decreased cell permeability or production of efflux pumps, (2) altering the antibiotic cellular target of the antibiotic through mutation, increased expression, or production of protection proteins, and (3) degrading the antibiotic molecule by production of antibiotic-inactivating enzymes [112, 113]. The resistome is the sum total of all AR genes in a microbiome [114, 115].

AR can lead to primary treatment failure, requiring the use of reserve or last resort antibiotics, whose use should be withheld except for the treatment of MDR infections that have exhausted all alternatives, such as the carbapenems, colistins, and third-generation tetracyclines [116, 117]. However, even highly-resistant strains can continue to acquire AR genes, including those against antibiotics of last resort for which there are limited treatment options [12, 18]. Often these drugs were initially abandoned due to unacceptable toxicities, but have since been revisited because they are the only remaining options. Plasmid-mediated resistance determinants against reserve antibiotics are significant

AR threats, particularly the bla_{NDM1} and bla_{KPC} carbapenemases [118–121] and QNR quinolone resistance proteins [122–124]. On the other hand, some AR genes, such as the *tet*(X)-like tetracycline destructases [125] and *mcr-1* colistin resistance gene [126], are similarly globally-distributed and capable of causing resistance against reserve antibiotics, but are rarely observed in clinical samples. While not an immediate threat, these genes' potential for broad dissemination into human pathogens warrants special attention and highlights the importance of surveilling environmental and human commensal metagenomes for novel AR genes [127].

Genotypic resistance can be determined using the same WGS and WMS data used to determine taxa, and declining sequencing costs have coincided with an ever-increasing catalog of known resistance genes [113]. A survey of resistance proteins in the UniProt database [128] showed that since 1986 there has been an exponential increase in the number of resistance determinants classified as β-lactamases, chloramphenicol acetyltransferases or tetracycline efflux pumps [129]. In recent years, studies of the resistome have expanded from single isolates to entire microbial communities, shedding light on the importance of environmental microbiomes as a reservoir for diverse genotypes [130, 131]. Additionally, recently-developed rapid diagnostic tools, which infer resistance based on growth phenotypes, promise to accelerate the timeline for when clinical microbiologists and physicians can determine the resistance profile of an HAI, and tailor treatment accordingly [80, 132, 133]. Here we highlight some of the main methods for investigating the resistome and examples of their application in the clinic.

Traditional methods for AR detection

Determining the resistances and susceptibilities of the clinical isolate underlying a HAI can refine treatment from broad-spectrum antibiotics—which indiscriminately kill the "good" commensal bacteria along with the pathogen, selecting for resistance and potentially leading to dysbiosis—to targeted, narrow-spectrum antibiotics. AR has traditionally been assayed by ASTs, which are culture-based assays that determine the concentration of antibiotic required to inhibit the growth of an isolate [134]. ASTs can be performed on solid media using Kirby-Bauer disks and gradient-diffusion strips, or in liquid media using microbroth dilutions. These data are then compared to standards (aka clinical breakpoints) published by the Clinical and Laboratory Standards Institute (CLSI) [135] and the European Committee on AST (EUCAST) [136] to be translated into resistance categories (resistant, intermediate, or susceptible). These assays continue to be widely used in hospital clinical microbiology laboratories because they provide actionable data to guide patient treatment decisions. However, ASTs have several limitations which make them unsuited for large-scale

surveillance of antibiotic resistance. First, they require cultured isolates, precluding insight into the resistome of the overall community. Second, they rely on published standards to interpret the results, which may be incomplete for the antibiotic or species in question, may utilize outdated methodology, or may conflict with the guidelines of other agencies [137]. Lastly, AST have low-resolution on the genetic underpinnings of resistance; however, screening of environmental or agricultural isolates using ASTs followed by WGS has led to the identification of many novel AR genes [138, 139].

A more high-throughput method of AR determination is multiplexed PCR. Panels of DNA primers and probes target and amplify known AR gene markers for numerous drug classes including methicillin [140], carbapenems [141], and tetracyclines [142]. Further, diagnostic kits such as the TaqMan Array Card (TAC) [143] target both AR genes and taxa-specific sequences, enabling the simultaneous detection of pathogens and specific AR genes. The quick turnaround rate of this approach enables rapid diagnosis of patients, and it is effective with sample types that have a low abundance of bacterial DNA and high abundance of contaminating human DNA (e.g., blood), which next-generation sequencing technologies have difficulty with. However, multiplexed PCR can only detect AR genes for which it has primers, ultimately making a comprehensive analysis of an entire resistome unfeasible.

NGS-based identification and discovery of AR genes

Analyzing WGS and WMS data for the presence of AR genes can predict functional features of bacteria without the laborious effort of culturing isolates and performing and interpreting ASTs or multiplexed PCR reactions. (Sequencing-based methods for AR gene detection are comprehensively reviewed in [113].) Assembly-based tools directly identify AR genes in assembled genomes, generated using the WGS and WMS tools described in the previous section, by predicting protein-coding regions and comparing them with AR gene databases. Resfinder [144] uses BLAST-based methods to identify resistance genes from bacterial isolate genomes, while Pointfinder [145] detects chromosomal point mutations associated with antibiotic resistance, and the Resistance Gene Identifier (RGI) [146] uses both pairwise comparisons and curated AR detection models to annotate AR genes. An additional bonus of assembly-based tools is the ability to determine the genetic contexts of AR genes by analyzing up- and downstream elements. These can include mobile elements which can facilitate the intra- or intercellular transfer of the AR gene, providing greater evidence for claims of horizontal transfer (described in more detail in a later section). Read-based tools, on the other hand, do not rely on prior assembly and can detect AR genes by directly aligning reads against reference databases, or splitting reads

into k-mers (subsequences of a length k) prior to a database search. SRST2 [147] and Kmer Resistance [148] are examples of each, respectively. These tools can identify AR genes in contaminated samples as well as low-abundance organisms, which would be unfit for de novo genome assembly. Though read-based tools lose contextual information about the position of genes within a genome and are prone to reporting false-positives due to sequence homology, by skipping de novo assembly and protein-coding prediction, they are typically faster and less computationally demanding than assembly-based tools.

As both methods are reliant on mapping to AR gene databases, it is important that the database used is up-to-date and appropriate for the sample type. Some AR gene databases are general, such as CARD (Comprehensive Antibiotic Resistance Database) [146], while others are specialized towards specific AR gene families (e.g., LacED (Lactamase Engineering Database) [149] and the Tuberculosis Drug Resistance Database [150]). Using multiple databases can mitigate some of the biases inherent to a single database, but this approach will still be restricted to the identification of known AR genes. However, a potent counterbalance to this limitation is the ability to re-analyze sequencing datasets or perform targeted PCR on stored isolates when a new AR gene is discovered.

For example, the discovery of the plasmid-mediated *mcr-1* colistin resistance gene [138] was significant because resistance to polymixins (an antibiotic of last resort) thus far had only involved chromosomal mutations. *mcr-1*, therefore, opened the door for rapid global dissemination of resistance against this antibiotic of last resort. The authors retrospectively identified *mcr-1* in 16/1,322 (1%) Enterobacteriaceae clinical isolates from two Chinese hospitals, but the full extent of its spread, both taxonomically and geographically, was unknown. Within months of that initial report, multiple retrospective analyses of clinical and environmental isolates using WGS and PCR were published in rapid succession, which collectively showed that the gene had already spread, undetected, into several pathogenic species cultured in many other countries around the world [151–157]. Now, *mcr-1* is a staple of AR gene databases, allowing it to be easily identified in WGS and WMS studies.

Several studies have aimed to characterize the resistomes of complex metagenomic samples, such as the human microbiome. However, referencing a large database of full-length AR genes can be computationally demanding for these complex metagenomic datasets, and the strict cutoffs for pairwise alignments can miss remote homologs and novel AR genes. As an alternative approach, ShortBRED (Short, Better Representative Extract Dataset) [158] constructs markers, short peptide sequences conserved within a protein family but distinct from other families, for AR gene families. Reads are then mapped to these marker sets, which can determine both

the prevalence and abundance of AR genes in a bacterial community [131, 159]. Another method is to use a database built using profile hidden Markov models (HMM), such as Resfams [160], instead of full-length protein sequences. Profile HMMs are statistical models trained on multiple-sequence alignments of genes with a known function. These are widely used for protein annotation [161, 162], and the authors of ResFams showed it could identify 64% more AR genes in soil and human gut metagenomes than pairwise alignment with the CARD and ARDB (Antibiotic Resistance Genes Database) databases [160, 163].

While the above in silico methods can detect known AR genes, functional metagenomics is a sequence-unbiased approach which can characterize resistomes and identify novel AR genes using the selective power of ASTs, without the laborious effort of culturing and screening isolates [164]. Functional metagenomic libraries are created by extracting total metagenomic DNA from a sample, packaging DNA fragments into an expression vector, and transforming that library of vectors into a susceptible host (typically *E. coli*). The library can then be screened for resistance via plating on antibiotic-containing media (or in other conditions when assaying for different phenotypes), and AR genes are identified by sequencing the surviving clones. PARFuMS (Parallel Annotation and Reassembly of Functional Metagenomic Selections) [127] is a bioinformatic pipeline for assembling and annotating reads from functional metagenomic selections. More importantly, this method can identify AR genes present in unculturable bacteria which has led to the discovery of novel AR genes and mechanisms [165, 166] and antibiotic biomolecules [167]. While the reliance on out-of-host expression of a gene may generate false negatives, this can be overcome using other more genetically tractable hosts (e.g., *Bacteroides thetaiotaomicron*, *Streptomyces coelicolor*). The FARME (Functional Antibiotic Resistance Metagenomic Element) [168] and ResfinderFG [169] databases are comprised of sequences identified by functional metagenomics, but not represented in other databases built primarily from AR genes in clinical isolates.

Rapid diagnostic tools

While ASTs determine an isolate's functional response to an antibiotic, and WGS characterizes its genetic repertoire of AR genes, the time, technical expertise, and resources required limits their clinical utility. In response, several new approaches have been developed which promise phenotypic resistance/susceptibility data in mere hours instead of days. These can rapidly inform modifications to patient treatment, empowering healthcare providers to modify treatment away from broad-spectrum antibiotics which may fail due to resistance, towards narrow-spectrum drugs that are effective and decrease general selection for antibiotic

resistance in hospitals. In theory, all of these approaches are generalizable to any pathogen-antibiotic pair of interest, but more work is needed before they can be utilized by clinical labs. In addition to method-specific technical details, each method described below will require larger and higher quality databases, as well as integration into systems for automated sample processing and detection, and benchmarking against traditional assays.

RNA detection assays are based on the premises that (1) RNA transcripts are species-specific, and (2) antibiotic exposure triggers transcriptional responses in susceptible, but not resistant, strains [170, 171]. Further, as transcriptional responses are among the first changes that occur upon drug exposure, they can be detected long before traditional growth-based phenotypes. GoPhAST-R (combined genotypic and phenotypic AST through RNA detection) [132] is a multiplexed mRNA detection assay, which can distinguish between susceptible and resistant strains of three ESKAPEE pathogens (*K. pneumoniae*, *A. baumannii*, and *E. coli*) when exposed to antibiotics of three major classes (fluoroquinolones, aminoglycosides and carbapenems) in < 4 h. It can also detect transcripts of key genetic resistance determinants, such as carbapenemases, to simultaneously profile phenotypic and genotypic information about resistance.

Alternatively, instead of directly identifying specific antibiotic resistance determinants, genetic linkage between resistance elements and the rest of the genome can enable the inference of resistance phenotypes based on coarse strain typing alone. Genomic neighbor typing [80] is an approach which determines how related a clinical isolate is to reference genomes with a known phylogeny and phenotype, providing predictions on the isolate's resistances and susceptibilities. By matching nanopore long-read sequencing data against a database of genomes in real time, the RASE (resistance-associated sequence elements) algorithm identifies a given sample's closest relatives. The authors demonstrate that this method can differentiate between resistant and susceptible isolates of *S. pneumoniae* and *N. gonorrhoeae* within 10 min of starting sequencing, and within 4 h of beginning sample preparation on clinical metagenomic samples containing *S. pneumoniae*.

MALDI-TOF MS, which has been widely adopted for species identification, can also be modified to simultaneously detect antimicrobial resistance or susceptibility [133]. This can be done by several innovative approaches [25]. One approach is to detect antibiotic degradation, as during the activity of beta-lactamases. Hydrolysis of the beta-lactam molecule changes the mass of a sample, which can be detected by MALDI-TOF [172]. Another approach involves identifying biomarkers specific to resistant organisms. Biomarker signals associated with pathogens such as MRSA [173], KPC-producing *K. pneumonia* [174], and *B. fragilis cfiA* [175] have already been identified with high specificity,

offering a highly accurate assessment of a sample without any additional cost. Finally, media containing isotopically-labeled amino acids can be incubated with organisms and antibiotics. Resistant organisms will synthesize proteins with the labeled amino acids, which can be identified by their distinct spectra [176, 177].

Resistome analyses are typically performed downstream of taxonomic analyses (e.g., using the same WMS or WGS datasets) and as such will similarly be influenced by sample choice and questions being asked. For example, while both WGS and rapid diagnostic tools are performed on cultured isolates, the speed of the diagnostic tools comes at the expense of not being able to directly identify the genetic underpinnings of resistance features—inferring resistance, instead, from correlated features. This trade-off may be acceptable for clinical diagnoses where every hour counts, but AR surveillance efforts may be better suited with WGS which generate higher-resolution data. Additionally, functional metagenomics can comprehensively identify all AR genes in a metagenomic sample, including novel genes, but it requires specialized library preparation. This extra step may require more effort than desired if a study is only interested in examining the presence of specific clinically-important AR genes. But banked isolate samples and genomes can be re-analyzed using updated tools and AR gene databases in future analyses to identify features that were missed by previous analyses, as evidenced with the retrospective analyses for *mcr-1*.

Identifying associations with mobile elements

The AR gene repertoire of a strain, as determined by the methods described in the previous section, is not a static identity. In addition to the slow processes of mutation and natural selection, nosocomial pathogens can rapidly acquire additional, fully-functioning AR genes via horizontal gene transfer (HGT). Indeed, many clinically-important ARGs have emerged via HGT from non-pathogenic reservoirs [127, 178–180]. Finding high-identity ARGs in two or more phylogenetically distinct species suggests a recent HGT event, but this could also be explained by other factors such as functional optimization or lower mutation rate [181]. Claims of recent transfer are strengthened if it can be demonstrated that the ARGs are flanked by the same mobile genetic elements (MGE)—genetic features that facilitate DNA movement. MGEs can be divided into two major groups: features that facilitate intercellular (i.e. HGT) or intracellular DNA mobility. MGEs can introduce gain-of-function mutations by shuttling "passenger" proteins (e.g., ARGs) between bacteria or DNA elements. Alternatively, they can lead to adaptive loss-of-function mutations by inserting into and interrupting

genes (e.g., those causing antibiotic susceptibility) or their regulatory elements [182]. The combined activities of various types of MGEs in a microbiome—collectively referred to as the mobilome or mobile gene pool [183, 184]—can rapidly accelerate the evolution of multidrug resistant bacteria.

HGT (i.e. intercellular transfer) can occur by three primary mechanisms: conjugation by plasmids, transduction by bacteriophages, and transformation by uptake of extracellular DNA [109] (Fig. 1b). All of these are capable of shuttling protein sequences (e.g., ARGs) between cells; however, conjugation by plasmids is thought to be the most consequential mechanism for the horizontal transfer of ARGs between species [185]. This is because conjugation is explicitly designed for the spread of bacterial genes—unlike transduction, which is an unintended consequence of bacteriophage replication. Conjugation also protects the DNA during its transit between cells, whereas in transformation, naked DNA must survive the extracellular environment. By facilitating the transfer of multiple ARGs at the same time, conjugation potentially enables recipient bacteria to become multi-drug resistant with just one HGT event. Once inside the new cell, the ARG can be expressed on the MGE directly, or be integrated into the genome of its new host or that of another plasmid via intracellular MGEs, such as insertion sequences, transposons, and integrons (Fig. 1b).

Recent advancements in long-read technologies, hybrid assembly, and bioinformatic tools can resolve the genomic context of plasmids in isolates and metagenomes, advancing the study of HGT beyond in vitro laboratory studies and into working hospitals and clinics [186–189]. It is beyond the scope of this review to cover all MGEs for all species (MGEs specific to ESKAPEE pathogens have been thoroughly reviewed in [190]). Instead, we will briefly describe the genomic approaches and bioinformatic tools that can be used to identify MGEs in clinical samples, with a special focus on plasmids.

NGS-based methods for MGE identification

Plasmid typing can provide insights into the epidemiology of HAI outbreaks, such as whether ARGs are carried by diverse vectors or a single epidemic plasmid. Several schemes have been developed to classify plasmids based on differences in conserved "backbone" regions, principal among them being replicon typing and MOB typing. Replicon typing targets loci that encode the plasmid's replication machinery and classifies them into the major plasmid incompatibility (Inc) groups, and plasmid multi-locus sequence typing (pMLST) can further assign them into subtypes within the broader replicon type [191–193]. MOB typing, on the other hand, specifically targets the relaxase protein of transmissible plasmids [194]. While originally developed for multiplexed

PCR reactions, these schemes have been adapted for in silico detection using tools such as PlasmidFinder [195, 196]. However, considering the ease with which it is now possible to sequence and compare entire plasmid genomes, this type of plasmid identification based only on small fragments has fallen out of favor, and these schemes are now used primarily for categorization [197, 198]. The principal limitation of replicon and MOB typing is that they are limited to known families, and neither can classify all plasmids—even among well-studied taxa. According to a recent estimate, 85% and 65% of publicly available plasmids can be replicon and MOB typed, respectively [199], though this is likely to be biased towards clinically-relevant plasmids with the highest representation in both plasmid sequence databases and typing schemes [200].

Identification of novel plasmids requires de novo assembly of WGS data. However, this task is complicated by the fact that it often is not clear which contigs belong to the chromosome and which to plasmids. By taking advantage of the fact that many plasmids exist at a higher copy number than the chromosome, PlasmidSPAdes [201] reconstructs plasmids from isolate WGS data using differences in coverage. For metagenomes, where differences in coverage are confounded by differences in relative abundance, several specialized tools for assembling plasmid sequences or identifying plasmid-associated contigs have been developed, including cBar [202], PlasFlow [203], Recycler [204], and metaplasmidSPAdes [200, 205]. In general, these tools can accurately predict small plasmids, especially those with well-known backbone structures, but have difficulty predicting large plasmids (> 50 kbp) [200]. As with chromosome assembly, large repetitive elements make it is nearly impossible to correctly assemble a plasmid genome using short-read WGS data alone, but this can be addressed by manually closing the genome using PCR.

To mitigate this limitation of WGS assembly, binning plasmid sequences from WGS data can be assisted by reference-based read mapping. By mapping reads against a closed reference (or index) plasmid, the presence of the reference can be inferred if there is sufficient homology across a given length (e.g., 99% sequence identity over 80% of the plasmid) [206, 207]. Reference plasmid sequences are typically assembled using long-read sequencing data and can pulled from the growing plasmid reference content in databases [208], or from a select sample, such as from the onset of an outbreak. This method has been successfully used to track patterns of introduction and spread of plasmids during outbreaks. For instance, one study of carbapenem-resistant *K. pneumoniae* (KPC) isolates from individuals from a Nepali neonatal unit used dataset-specific reference sequences to determine a key strain containing four highly conserved plasmids (including one carrying $bla_{\text{NDM-1}}$) which was responsible for multiple clusters of cases [209]. This

approach, however, assumes that the index plasmid is conserved throughout the study, and that plasmid structures are largely conserved. However, this may not be the case as plasmid structure can be extremely dynamic over time—even within short-term outbreaks—with smaller MGEs facilitating structural rearrangements through homologous recombination with other plasmids, or by adding/losing copies of genes. Precisely because these types of repetitive elements are difficult for short-read data to resolve, reference-based approaches can lead to the incorrect conclusion that the reconstructed plasmid is identical to the reference, when in reality significant rearrangements have occurred [207].

As previously described, these issues with de novo assembly can be largely overcome using long-read sequencing technologies, such as PacBio and nanopore sequencing [48, 49]. For example, in a study of carbapenem resistance in hospital-derived *Enterobacteriaceae*, assembly with Illumina data indicated that 11/17 sequenced isolates were predicted to carry the bla_{KPC} gene on a pKPC-UVA01-like plasmid. This was expected as pKPC-UVA01 is a known index for bla_{KPC} presence, and suggested that the dominant mechanism for dissemination of the bla_{KPC} gene was via carriage in pKPC-UVA01. However, while PacBio assemblies confirmed that all 11 isolates indeed contained a pKPC-UVA01-like plasmid, only 5/11 of those plasmids contained bla_{KPC} (i.e. were identical to the index). In the other 6/11 isolates, other plasmids—some only present only one of them—carried bla_{KPC}, showing that the gene's dissemination is actually driven by several distinct vectors and homologous recombination between them [207].

Long-read technologies are also useful when the goal is to generate contiguous sequences of novel plasmids [210]. For instance, the fact that PacBio read lengths can be significantly longer than plasmid genomes can be taken advantage of to increase coverage through an approach called circular consensus sequencing (CCS) [211], where a single read produces multiple observations of each base, thereby increasing accuracy. PacBio sequencing was used to assemble the sequences of megaplasmids (> 420 kb) in a study of MDR *P. aeruginosa* isolates from a hospital in Thailand [212], and these assemblies showed that the resistance regions of these megaplasmids harbored multiple ARGs and were rich with smaller MGEs. Nanopore sequencing has also been coupled with culture-based enrichment of metagenomic samples swabbed from hospital surfaces, which resulted in the generation of 5,910 closed plasmid sequences, many > 100 kb long [19].

While useful on their own, these methods are most effective when coupled together; WGS with reference-based read mapping can identify interesting isolates (e.g., those containing plasmids with ARGs of interest or likely HGT events), which can inform decisions about which isolates are chosen for long-read sequencing. Then the resulting contiguous

plasmid sequences can be typed using replicon or MOB schemes. This approach was taken in an 18-month study of 2,173 clinical isolates from a single hospital [181], where short-read WGS data identified high-identity sequences shared by different genera, and then nanopore sequencing and hybrid assembly was performed on a subset of isolates containing these shared sequences. This revealed a diverse group of high-identity ARG-encoding plasmids shared within and across species and genera. To take this a step further, the authors matched high-identity plasmids with patient metadata to identify epidemiological links, revealing instances of putative HGT between different genera within and between patients. For example, isolates of *K. pneumoniae* and *E. coli* collected on the same date from the same site from Patient A each carried an identical (99.95% identity) 113.6 kb beta-lactamase-carrying IncF plasmid—strong evidence for within-patient transfer. This same plasmid (99.98% identity) was later identified in an *E. coli* isolate of another species type cultured from Patient B 109 days after Patient A's sampling—strongly suggesting between-patient transfer. Further, Patient B had an adjacent hospital room to Patient A's and was treated by the same healthcare staff, strongly suggesting a hospital-associated reservoir and transmission network, though the connecting links were not able to be identified because surfaces were not examined.

Lastly, one step deeper from plasmid sequencing and identification has been that of the smaller intracellular MGEs—including insertion sequences (IS), transposons, and integrons—that can be nested within plasmid genomes. These can contribute to the spread of resistance by facilitating the transfer of AR genes from one genetic element to another, or interrupting susceptibility genes [182, 207]. In a process analogous to identifying AR genes from short-read sequencing data, MGEs can be identified by mapping to reads or contigs to databases of known elements or HMM profiles using tools such as ISfinder [213], ISMapper [214], and IntegronFinder [215]. However, identifying novel elements using short-read technologies is difficult because these elements are repetitive and often longer than a single read length. Structural variant tools, such as panISa [216] and MGEFinder [217], can identify novel insertion sequences by aligning a draft genome assembly to a reference genome of the same species (98.5% identity), and then identifies reads that have "clipped" (unaligned) ends—indicating the genomic site where an insertion sequence begins or ends.

Plasmid-specific approaches

Plasmid identification from WMS datasets has difficulties classifying metagenomic sequencing reads and contigs as plasmid- or chromosome-derived, confounding efforts to assemble novel plasmids; however, these technical difficulties can be circumvented by first isolating plasmids prior

to sequencing. Here, the so-called plasmidome or metamobilome is separated from chromosomes and other genetic elements by enzymatic digestion of sheared DNA—the vast majority of which is chromosomal [183]. Then, using transposon-aided capture (TRACA), these purified plasmids are captured by inserting a transposon containing an origin of replication and a selectable marker, and then transformed into an *E. coli* host and sequenced [218]. Alternatively, the purified plasmids can be amplified using inverse-PCR coupled with multiple displacement amplification prior to sequencing [218–220]. The principal limitations of these approaches are that they require additional sample preparation steps. Plasmid isolation methods are biased towards smaller plasmids (3–10 kb) because the digestion process can degrade large plasmids which were unintentionally sheared during the extraction process. Lastly, as with all assemblies using WGS data, manual PCR or long-read sequencing is required to fully close gaps [186].

While plasmid purification can produce hundreds of circular novel plasmids, it and WMS generally often cannot assign plasmids to their bacterial hosts because the DNA extraction process separates plasmids from chromosomes. Proximity-ligation methods, such as Hi-C and 3C, were originally developed to study three-dimensional genome structure in eukaryotic cells [221] but have recently been adapted to detect interactions between DNA molecules in bacterial metagenomes. This is accomplished by cross-linking DNA in close physical proximity—such as a plasmid and chromosome from the same cell—followed by proximity ligation of cross-linked DNA and short-read sequencing. Linkages between non-contiguous DNA can then be used by specialized clustering algorithms [222, 223] to deconvolute which plasmids and chromosomes originated from the same cell. As with other metagenomic approaches, Hi-C's resolution is limited by sample complexity and sequencing depth, which is further complicated by the fact that the majority of read pairs do not contain proximity ligation junctions (i.e. are not Hi-C read pairs). For example, the studies of human fecal samples estimated the fraction of Hi-C read pairs was between 0.36–0.67% and 1.38–2.38% [223, 224]. This will necessitate relatively deep sequencing (~ 40 to 70 million reads), likely limiting this approach to a small subset of samples.

Much work is still needed to optimize and scale-up Hi-C methodology for use in large surveillance efforts. In addition to testing with simulated and synthetic metagenomic samples [222, 225–227], it has been applied to single fecal samples from a healthy human and cow [223, 224, 228], used to compare fecal samples collected 10 years apart from two individuals [229], and to study plasmid-host associations in wastewater communities [230]. To our knowledge, just one has used this technique to analyze metagenomic samples from hospitalized patients [231]. Here, seven neutropenic patients undergoing hematopoietic stem cell transplantation were compared to two healthy individuals using samples from multiple timepoints over a 2–3 week period. This study found that networks of HGT were unique to each individual, and that while healthy patients had a basal level of HGT this was elevated in the neutropenic patients [231]. This suggests that these at-risk patients are at an increased risk for HGT of AR genes into MDR pathogens, and reinforces the role of patients' gut microbiomes as a reservoir of AR genes in the hospital.

Recent advancements in NGS technologies and bioinformatic tools have empowered researchers to study the spread of AR on plasmids and other MGEs at a resolution not possible even just a few years ago. Studies using these technologies have revealed a second layer of transmission during HAI outbreaks—of genetic transfer of AR genes between genetic elements and bacterial strains, instead of geographic transfer of bacterial strains between patients and/or the hospital environment. Comparative genomic analyses can be done on just those MGEs found up- and downstream of AR genes, or by resolving and tracking entire plasmids in a manner similar to tracking strains. Given the importance of HGT to the spread of novel resistance genes into and between HAI-causing pathogens, analyses of MGEs are sure to become a staple of outbreak tracking.

Conclusion

Despite extensive efforts to control infection, hospital environments are reservoirs of an incredible diversity of AR and MDR bacteria. New approaches have enabled (1) the identification of species and tracking of strains, (2) the rapid profiling of genotypic and phenotypic resistance, and (3) the resolution of MGEs that facilitate intra- and intercellular gene transfer, at a greater resolution than ever before. This data, coupled with longitudinal sampling and comparative genomics, has led significant insights into the spatiotemporal dynamics of hospital-associated bacterial communities and their AR gene cargo. This has revealed that the hospital microbiome is made up of complex nested systems, with multi-layered transmission of strains, plasmids, and smaller genetic elements between patients, healthcare workers, and hospital surfaces. In many cases, the same sequencing datasets can be re-analyzed with different bioinformatic tools to answer each of these questions, but often the most relevant data can only be generated with question-specific methods and unique sample preparation schemes. The approaches and tools used should be tailored to the goals of the study (e.g., complete environmental characterization, or detailed study of ESKAPE pathogens), and the sample types and technical expertise available. Continued research and technological advances are needed before these approaches can be

routinely applied for hospital surveillance efforts, but there is much promise in their ability to track outbreaks of AR and MDR bacteria, identify persistent environmental reservoirs, and gauge future risks.

References

1. Poe EA (1839) The Fall of the House of Usher. https://americanenglish.state.gov/files/ae/resource_files/the_fall_of_the_house_of_usher.pdf

2. Haque M, Sartelli M, McKimm J, Abu Bakar M (2018) Health care-associated infections—an overview. Infect Drug Resist 11:2321–2333. https://doi.org/10.2147/IDR.S177247

3. Lax S, Gilbert JA (2015) Hospital-associated microbiota and implications for nosocomial infections. Trends Mol Med 21(7):427–432. https://doi.org/10.1016/j.molmed.2015.03.005

4. Magill SS, Edwards JR, Bamberg W, Beldavs ZG, Dumyati G, Kainer MA, Lynfield R, Maloney M, McAllister-Hollod L, Nadle J, Ray SM, Thompson DL, Wilson LE, Fridkin SK, Healthcare-Associated EIP, I, Antimicrobial Use Prevalence Survey T, (2014) Multistate point-prevalence survey of health care-associated infections. N Engl J Med 370(13):1198–1208. https://doi.org/10.1056/NEJMoa1306801

5. Mora M, Mahnert A, Koskinen K, Pausan MR, Oberauner-Wappis L, Krause R, Perras AK, Gorkiewicz G, Berg G, Moissl-Eichinger C (2016) Microorganisms in confined habitats: microbial monitoring and control of intensive care units, operating rooms, cleanrooms and the international space station. Front Microbiol 7:1573. https://doi.org/10.3389/fmicb.2016.01573

6. Weiner LM, Webb AK, Limbago B, Dudeck MA, Patel J, Kallen AJ, Edwards JR, Sievert DM (2016) Antimicrobial-resistant pathogens associated with healthcare-associated infections: summary of data reported to the national healthcare safety network at the centers for disease control and prevention, 2011–2014. Infect Control Hosp Epidemiol 37(11):1288–1301. https://doi.org/10.1017/ice.2016.174

7. Hidron AI, Edwards JR, Patel J, Horan TC, Sievert DM, Pollock DA, Fridkin SK, National Healthcare Safety Network T, Participating National Healthcare Safety Network F (2008) NHSN annual update: antimicrobial-resistant pathogens associated with healthcare-associated infections: annual summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2006–2007. Infect Control Hosp Epidemiol 29(11):996–1011. https://doi.org/10.1086/591861

8. Vincent JL, Sakr Y, Singer M, Martin-Loeches I, Machado FR, Marshall JC, Finfer S, Pelosi P, Brazzi L, Aditianingsih D, Timsit JF, Du B, Wittebole X, Maca J, Kannan S, Gorordo-Delsol LA, De Waele JJ, Mehta Y, Bonten MJM, Khanna AK, Kollef M, Human M, Angus DC, Investigators EI (2020) Prevalence and outcomes of infection among patients in intensive care units in 2017. JAMA. https://doi.org/10.1001/jama.2020.2717

9. Weiner-Lastinger LM, Abner S, Edwards JR, Kallen AJ, Karlsson M, Magill SS, Pollock D, See I, Soe MM, Walters MS, Dudeck MA (2020) Antimicrobial-resistant pathogens associated with adult healthcare-associated infections: summary of data reported to the National Healthcare Safety Network, 2015–2017. Infect Control Hosp Epidemiol 41(1):1–18. https://doi.org/10.1017/ice.2019.296

10. Weiner-Lastinger LM, Abner S, Benin AL, Edwards JR, Kallen AJ, Karlsson M, Magill SS, Pollock D, See I, Soe MM, Walters MS, Dudeck MA (2020) Antimicrobial-resistant pathogens associated with pediatric healthcare-associated infections: summary of data reported to the National Healthcare Safety Network, 2015–2017. Infect Control Hosp Epidemiol 41(1):19–30. https://doi.org/10.1017/ice.2019.297

11. Berendonk TU, Manaia CM, Merlin C, Fatta-Kassinos D, Cytryn E, Walsh F, Burgmann H, Sorum H, Norstrom M, Pons MN, Kreuzinger N, Huovinen P, Stefani S, Schwartz T, Kisand V, Baquero F, Martinez JL (2015) Tackling antibiotic resistance: the environmental framework. Nat Rev Microbiol 13(5):310–317. https://doi.org/10.1038/nrmicro3439

12. CDC (2019) Antibiotic resistance threats in the United States, 2019. Atlanta, GA

13. Cassini A, Hogberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS, Colomb-Cotinat M, Kretzschmar ME, Devleesschauwer B, Cecchini M, Ouakrim DA, Oliveira TC, Struelens MJ, Suetens C, Monnet DL, Burden of AMRCG (2019) 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(1):56–66. https://doi.org/10.1016/S1473-3099(18)30605-4

14. Bush K, Courvalin P, Dantas G, Davies J, Eisenstein B, Huovinen P, Jacoby GA, Kishony R, Kreiswirth BN, Kutter E, Lerner SA, Levy S, Lewis K, Lomovskaya O, Miller JH, Mobashery S, Piddock LJ, Projan S, Thomas CM, Tomasz A, Tulkens PM, Walsh TR, Watson JD, Witkowski J, Witte W, Wright G, Yeh P, Zgurskaya HI (2011) Tackling antibiotic resistance. Nat Rev Microbiol 9(12):894–896. https://doi.org/10.1038/nrmicro2693

15. De Oliveira DMP, Forde BM, Kidd TJ, Harris PNA, Schembri MA, Beatson SA, Paterson DL, Walker MJ (2020) Antimicrobial resistance in ESKAPE pathogens. Clin Microbiol Rev. https://doi.org/10.1128/CMR.00181-19

16. Rice LB (2008) Federal funding for the study of antimicrobial resistance in nosocomial pathogens: no ESKAPE. J Infect Dis 197(8):1079–1081. https://doi.org/10.1086/533452

17. Mulani MS, Kamble EE, Kumkar SN, Tawre MS, Pardesi KR (2019) Emerging strategies to combat ESKAPE pathogens in the Era of antimicrobial resistance: a review. Front Microbiol 10:539. https://doi.org/10.3389/fmicb.2019.00539

18. Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, Pulcini C, Kahlmeter G, Kluytmans J, Carmeli Y, Ouellette M, Outterson K, Patel J, Cavaleri M, Cox EM, Houchens CR, Grayson ML, Hansen P, Singh N, Theuretzbacher U, Magrini N, Group WHOPPLW (2018) Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. Lancet Infect Dis 18(3):318–327. https://doi.org/10.1016/S1473-3099(17)30753-3

19. Chng KR, Li C, Bertrand D, Ng AHQ, Kwah JS, Low HM, Tong C, Natrajan M, Zhang MH, Xu L, Ko KKK, Ho EXP, Av-Shalom

TV, Teo JWP, Khor CC, Meta SUBC, Chen SL, Mason CE, Ng OT, Marimuthu K, Ang B, Nagarajan N (2020) Cartography of opportunistic pathogens and antibiotic resistance genes in a tertiary hospital environment. Nat Med 26(6):941–951. https://doi.org/10.1038/s41591-020-0894-4

20. Weingarten RA, Johnson RC, Conlan S, Ramsburg AM, Dekker JP, Lau AF, Khil P, Odom RT, Deming C, Park M, Thomas PJ, Program NCS, Henderson DK, Palmore TN, Segre JA, Frank KM (2018) Genomic analysis of hospital plumbing reveals diverse reservoir of bacterial plasmids conferring carbapenem resistance. mBio. https://doi.org/10.1128/mBio.02011-17

21. D'Souza AW, Potter RF, Wallace M, Shupe A, Patel S, Sun X, Gul D, Kwon JH, Andleeb S, Burnham CD, Dantas G (2019) Spatiotemporal dynamics of multidrug resistant bacteria on intensive care unit surfaces. Nat Commun 10(1):4569. https://doi.org/10.1038/s41467-019-12563-1

22. Hu H, Johani K, Gosbell IB, Jacombs AS, Almatroudi A, Whiteley GS, Deva AK, Jensen S, Vickery K (2015) Intensive care unit environmental surfaces are contaminated by multidrug-resistant bacteria in biofilms: combined results of conventional culture, pyrosequencing, scanning electron microscopy, and confocal laser microscopy. J Hosp Infect 91(1):35–44. https://doi.org/10.1016/j.jhin.2015.05.016

23. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239(4839):487–491. https://doi.org/10.1126/science.2448875

24. Buchan BW, Ledeboer NA (2014) Emerging technologies for the clinical microbiology laboratory. Clin Microbiol Rev 27(4):783–822. https://doi.org/10.1128/CMR.00003-14

25. Hou TY, Chiang-Ni C, Teng SH (2019) Current status of MALDI-TOF mass spectrometry in clinical microbiology. J Food Drug Anal 27(2):404–414. https://doi.org/10.1016/j.jfda.2019.01.001

26. Balada-Llasat JM, Kamboj K, Pancholi P (2013) Identification of mycobacteria from solid and liquid media by matrix-assisted laser desorption ionization-time of flight mass spectrometry in the clinical laboratory. J Clin Microbiol 51(9):2875–2879. https://doi.org/10.1128/JCM.00819-13

27. Mesureur J, Arend S, Celliere B, Courault P, Cotte-Pattat PJ, Totty H, Deol P, Mick V, Girard V, Touchberry J, Burrowes V, Lavigne JP, O'Callaghan D, Monnin V, Keriel A (2018) A MALDI-TOF MS database with broad genus coverage for species-level identification of Brucella. PLoS Negl Trop Dis 12(10):e0006874. https://doi.org/10.1371/journal.pntd.0006874

28. Seuylemezian A, Aronson HS, Tan J, Lin M, Schubert W, Vaishampayan P (2018) Development of a custom MALDI-TOF MS database for species-level identification of bacterial isolates collected from spacecraft and associated surfaces. Front Microbiol 9:780. https://doi.org/10.3389/fmicb.2018.00780

29. Veloo ACM, Jean-Pierre H, Justesen US, Morris T, Urban E, Wybo I, Kostrzewa M, Friedrich AW, workgroup E (2018) Validation of MALDI-TOF MS Biotyper database optimized for anaerobic bacteria: the ENRIA project. Anaerobe 54:224–230. https://doi.org/10.1016/j.anaerobe.2018.03.007

30. Wetterstrand KA (2019) DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP). https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data.

31. Quainoo S, Coolen JPM, van Hijum S, Huynen MA, Melchers WJG, van Schaik W, Wertheim HFL (2017) Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. Clin Microbiol Rev 30(4):1015–1063. https://doi.org/10.1128/CMR.00016-17

32. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW (2012) Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet 13(9):601–612. https://doi.org/10.1038/nrg3226

33. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18(5):821–829. https://doi.org/10.1101/gr.074492.107

34. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5):455–477. https://doi.org/10.1089/cmb.2012.0021

35. Wick RR, Judd LM, Gorrie CL, Holt KE (2017) Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13(6):e1005595. https://doi.org/10.1371/journal.pcbi.1005595

36. Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29(8):1072–1075. https://doi.org/10.1093/bioinformatics/btt086

37. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM (2016) Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol 17(1):132. https://doi.org/10.1186/s13059-016-0997-x

38. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35(9):3100–3108. https://doi.org/10.1093/nar/gkm160

39. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, Chun J (2017) Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 67(5):1613–1617. https://doi.org/10.1099/ijsem.0.001755

40. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30(14):2068–2069. https://doi.org/10.1093/bioinformatics/btu153

41. Douglas AP, Marshall C, Baines SL, Ritchie D, Szer J, Madigan V, Chan HT, Ballard SA, Howden BP, Buising K, Slavin MA (2019) 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(3):303–308. https://doi.org/10.1099/jmm.0.000916

42. Johnson RC, Deming C, Conlan S, Zellmer CJ, Michelin AV, Lee-Lin S, Thomas PJ, Park M, Weingarten RA, Less J, Dekker JP, Frank KM, Musser KA, McQuiston JR, Henderson DK, Lau AF, Palmore TN, Segre JA (2018) Investigation of a cluster of sphingomonas koreensis infections. N Engl J Med 379(26):2529–2539. https://doi.org/10.1056/NEJMoa1803238

43. Decraene V, Phan HTT, George R, Wyllie DH, Akinremi O, Aiken Z, Cleary P, Dodgson A, Pankhurst L, Crook DW, Lenney C, Walker AS, Woodford N, Sebra R, Fath-Ordoubadi F, Mathers AJ, Seale AC, Guiver M, McEwan A, Watts V, Welfare W, Stoesser N, Cawthorne J, Group TI (2018) 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. https://doi.org/10.1128/AAC.01689-18

44. Maatallah M, Vading M, Kabir MH, Bakhrouf A, Kalin M, Naucler P, Brisse S, Giske CG (2014) 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(11):e113539. https://doi.org/10.1371/journal.pone.0113539

45. Long SW, Linson SE, Ojeda Saavedra M, Cantu C, Davis JJ, Brettin T, Olsen RJ (2017) Whole-genome sequencing of human clinical klebsiella pneumoniae isolates reveals misidentification and misunderstandings of Klebsiella pneumoniae,

Klebsiella variicola, and Klebsiella quasipneumoniae. mSphere. https://doi.org/10.1128/mSphereDirect.00290-17

46. Berry GJ, Loeffelholz MJ, Williams-Bouyer N (2015) An investigation into laboratory misidentification of a bloodstream Klebsiella variicola infection. J Clin Microbiol 53(8):2793–2794. https://doi.org/10.1128/JCM.00841-15

47. Potter RF, Lainhart W, Twentyman J, Wallace MA, Wang B, Burnham CA, Rosen DA, Dantas G (2018) Population structure, antibiotic resistance, and uropathogenicity of Klebsiella variicola. mBio. https://doi.org/10.1128/mbio.02481-18

48. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S (2009) Real-time DNA sequencing from single polymerase molecules. Science 323(5910):133–138. https://doi.org/10.1126/science.1162986

49. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H (2009) Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol 4(4):265–270. https://doi.org/10.1038/nnano.2009.12

50. Mikheyev AS, Tin MM (2014) A first look at the Oxford Nanopore MinION sequencer. Mol Ecol Resour 14(6):1097–1102. https://doi.org/10.1111/1755-0998.12324

51. Magi A, Giusti B, Tattini L (2017) Characterization of MinION nanopore data for resequencing analyses. Brief Bioinform 18(6):940–953. https://doi.org/10.1093/bib/bbw077

52. Magi A, Semeraro R, Mingrino A, Giusti B, D'Aurizio R (2018) Nanopore sequencing data analysis: state of the art, applications and challenges. Brief Bioinform 19(6):1256–1272. https://doi.org/10.1093/bib/bbx062

53. Antipov D, Korobeynikov A, McLean JS, Pevzner PA (2016) hybridSPAdes: an algorithm for hybrid assembly of short and long reads. Bioinformatics 32(7):1009–1015. https://doi.org/10.1093/bioinformatics/btv688

54. Utturkar SM, Klingeman DM, Land ML, Schadt CW, Doktycz MJ, Pelletier DA, Brown SD (2014) Evaluation and validation of de novo and hybrid assembly techniques to derive high-quality genome sequences. Bioinformatics 30(19):2709–2716. https://doi.org/10.1093/bioinformatics/btu391

55. Coenye T, Vandamme P (2003) Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. FEMS Microbiol Lett 228(1):45–49. https://doi.org/10.1016/S0378-1097(03)00717-1

56. Weisburg WG, Barns SM, Pelletier DA, Lane DJ (1991) 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol 173(2):697–703. https://doi.org/10.1128/jb.173.2.697-703.1991

57. Rampelotto PH, Sereia AFR, de Oliveira LFV, Margis R (2019) Exploring the Hospital microbiome by high-resolution 16S rRNA profiling. Int J Mol Sci. https://doi.org/10.3390/ijms20123099

58. ElRakaiby MT, Gamal-Eldin S, Amin MA, Aziz RK (2019) Hospital Microbiome Variations As Analyzed by High-Throughput Sequencing. OMICS 23(9):426–438. https://doi.org/10.1089/omi.2019.0111

59. Liu Z, DeSantis TZ, Andersen GL, Knight R (2008) Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic Acids Res 36(18):e120. https://doi.org/10.1093/nar/gkn491

60. Schloss PD, Westcott SL (2011) Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. Appl Environ Microbiol 77(10):3219–3226. https://doi.org/10.1128/AEM.02810-10

61. Brooks JP, Edwards DJ, Harwich MD Jr, Rivera MC, Fettweis JM, Serrano MG, Reris RA, Sheth NU, Huang B, Girerd P, Vaginal Microbiome C, Strauss JF 3rd, Jefferson KK, Buck GA (2015) The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. BMC Microbiol 15:66. https://doi.org/10.1186/s12866-015-0351-6

62. Callahan BJ, McMurdie PJ, Holmes SP (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J 11(12):2639–2643. https://doi.org/10.1038/ismej.2017.119

63. Hewitt KM, Mannino FL, Gonzalez A, Chase JH, Caporaso JG, Knight R, Kelley ST (2013) Bacterial diversity in two Neonatal Intensive Care Units (NICUs). PLoS ONE 8(1):e54703. https://doi.org/10.1371/journal.pone.0054703

64. Brooks B, Firek BA, Miller CS, Sharon I, Thomas BC, Baker R, Morowitz MJ, Banfield JF (2014) Microbes in the neonatal intensive care unit resemble those found in the gut of premature infants. Microbiome 2(1):1. https://doi.org/10.1186/2049-2618-2-1

65. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, Knight R, Knights D (2018) Evaluating the information content of shallow shotgun metagenomics. mSystems. https://doi.org/10.1128/mSystems.00069-18

66. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N (2015) MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods 12(10):902–903. https://doi.org/10.1038/nmeth.3589

67. Wood DE, Salzberg SL (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15(3):R46. https://doi.org/10.1186/gb-2014-15-3-r46

68. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, Lipson KS, Knight R, Caporaso JG, Segata N, Huttenhower C (2018) Species-level functional profiling of metagenomes and metatranscriptomes. Nat Methods 15(11):962–968. https://doi.org/10.1038/s41592-018-0176-y

69. Tamburini FB, Andermann TM, Tkachenko E, Senchyna F, Banaei N, Bhatt AS (2018) Precision identification of diverse bloodstream pathogens in the gut microbiome. Nat Med 24(12):1809–1814. https://doi.org/10.1038/s41591-018-0202-8

70. Ghurye JS, Cepeda-Espinoza V, Pop M (2016) Metagenomic assembly: overview, challenges and applications. Yale J Biol Med 89(3):353–362

71. Li D, Liu CM, Luo R, Sadakane K, Lam TW (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31(10):1674–1676. https://doi.org/10.1093/bioinformatics/btv033

72. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA (2017) metaSPAdes: a new versatile metagenomic assembler. Genome Res 27(5):824–834. https://doi.org/10.1101/gr.213959.116

73. Namiki T, Hachiya T, Tanaka H, Sakakibara Y (2012) MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic Acids Res 40(20):e155. https://doi.org/10.1093/nar/gks678

74. Afiahayati SK, Sakakibara Y (2015) MetaVelvet-SL: an extension of the Velvet assembler to a de novo metagenomic assembler utilizing supervised learning. DNA Res 22(1):69–77. https://doi.org/10.1093/dnares/dsu041

75. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glockner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP,

Ettema TJG, Tighe S, Konstantinidis KT, Liu WT, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Genome Standards C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35(8):725–731. https://doi.org/10.1038/nbt.3893

76. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C (2014) Binning metagenomic contigs by coverage and composition. Nat Methods 11(11):1144–1146. https://doi.org/10.1038/nmeth.3103

77. Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC (2019) New insights from uncultivated genomes of the global human gut microbiome. Nature 568(7753):505–510. https://doi.org/10.1038/s41586-019-1058-x

78. Moss EL, Maghini DG, Bhatt AS (2020) Complete, closed bacterial genomes from microbiomes using nanopore sequencing. Nat Biotechnol 38(6):701–707. https://doi.org/10.1038/s41587-020-0422-6

79. Bertrand D, Shaw J, Kalathiyappan M, Ng AHQ, Kumar MS, Li C, Dvornicic M, Soldo JP, Koh JY, Tong C, Ng OT, Barkham T, Young B, Marimuthu K, Chng KR, Sikic M, Nagarajan N (2019) Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. Nat Biotechnol 37(8):937–944. https://doi.org/10.1038/s41587-019-0191-2

80. Brinda K, Callendrello A, Ma KC, MacFadden DR, Charalampous T, Lee RS, Cowley L, Wadsworth CB, Grad YH, Kucherov G, O'Grady J, Baym M, Hanage WP (2020) Rapid inference of antibiotic resistance and susceptibility by genomic neighbour typing. Nat Microbiol 5(3):455–464. https://doi.org/10.1038/s41564-019-0656-6

81. Kizny Gordon AE, Mathers AJ, Cheong EYL, Gottlieb T, Kotay S, Walker AS, Peto TEA, Crook DW, Stoesser N (2017) The hospital water environment as a reservoir for carbapenem-resistant organisms causing hospital-acquired infections-a systematic review of the literature. Clin Infect Dis 64(10):1435–1444. https://doi.org/10.1093/cid/cix132

82. Kanamori H, Weber DJ, Rutala WA (2016) Healthcare outbreaks associated with a water reservoir and infection prevention strategies. Clin Infect Dis 62(11):1423–1435. https://doi.org/10.1093/cid/ciw122

83. Ciufo S, Kannan S, Sharma S, Badretdin A, Clark K, Turner S, Brover S, Schoch CL, Kimchi A, DiCuccio M (2018) Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. Int J Syst Evol Microbiol 68(7):2386–2392. https://doi.org/10.1099/ijsem.0.002809

84. Brooks B, Olm MR, Firek BA, Baker R, Thomas BC, Morowitz MJ, Banfield JF (2017) Strain-resolved analysis of hospital rooms and infants reveals overlap between the human and room microbiome. Nat Commun 8(1):1814. https://doi.org/10.1038/s41467-017-02018-w

85. Segerman B (2012) The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. Front Cell Infect Microbiol 2:116. https://doi.org/10.3389/fcimb.2012.00116

86. Rokas A, Williams BL, King N, Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425(6960):798–804. https://doi.org/10.1038/nature02053

87. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J (2015) Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31(22):3691–3693. https://doi.org/10.1093/bioinformatics/btv421

88. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313. https://doi.org/10.1093/bioinformatics/btu033

89. Mostowy R, Croucher NJ, Andam CP, Corander J, Hanage WP, Marttinen P (2017) Efficient inference of recent and ancestral recombination within bacterial populations. Mol Biol Evol 34(5):1167–1182. https://doi.org/10.1093/molbev/msx066

90. Corander J, Waldmann P, Sillanpaa MJ (2003) Bayesian analysis of genetic differentiation between populations. Genetics 163(1):367–374

91. Dijkshoorn L, Ursing BM, Ursing JB (2000) Strain, clone and species: comments on three basic concepts of bacteriology. J Med Microbiol 49(5):397–401. https://doi.org/10.1099/0022-1317-49-5-397

92. Stimson J, Gardy J, Mathema B, Crudu V, Cohen T, Colijn C (2019) Beyond the SNP threshold: identifying outbreak clusters using inferred transmissions. Mol Biol Evol 36(3):587–603. https://doi.org/10.1093/molbev/msy242

93. Clark TG, Mallard K, Coll F, Preston M, Assefa S, Harris D, Ogwang S, Mumbowa F, Kirenga B, O'Sullivan DM, Okwera A, Eisenach KD, Joloba M, Bentley SD, Ellner JJ, Parkhill J, Jones-Lopez EC, McNerney R (2013) Elucidating emergence and transmission of multidrug-resistant tuberculosis in treatment experienced patients by whole genome sequencing. PLoS ONE 8(12):e83012. https://doi.org/10.1371/journal.pone.0083012

94. Coll F, Harrison EM, Toleman MS, Reuter S, Raven KE, Blane B, Palmer B, Kappeler ARM, Brown NM, Torok ME, Parkhill J, Peacock SJ (2017) Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community. Sci Transl Med. https://doi.org/10.1126/scitranslmed.aak9745

95. Silva IN, Santos PM, Santos MR, Zlosnik JE, Speert DP, Buskirk SW, Bruger EL, Waters CM, Cooper VS, Moreira LM (2016) Long-term evolution of burkholderia multivorans during a chronic cystic fibrosis infection reveals shifting forces of selection. mSystems. https://doi.org/10.1128/mSystems.00029-16

96. Zhao S, Lieberman TD, Poyet M, Kauffman KM, Gibbons SM, Groussin M, Xavier RJ, Alm EJ (2019) Adaptive evolution within gut microbiomes of healthy people. Cell Host Microbe 25(5):656-667 e658. https://doi.org/10.1016/j.chom.2019.03.007

97. Lieberman TD, Flett KB, Yelin I, Martin TR, McAdam AJ, Priebe GP, Kishony R (2014) Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. Nat Genet 46(1):82–87. https://doi.org/10.1038/ng.2848

98. David MD, Kearns AM, Gossain S, Ganner M, Holmes A (2006) Community-associated meticillin-resistant Staphylococcus aureus: nosocomial transmission in a neonatal unit. J Hosp Infect 64(3):244–250. https://doi.org/10.1016/j.jhin.2006.06.022

99. Rohde H, Qin J, Cui Y, Li D, Loman NJ, Hentschke M, Chen W, Pu F, Peng Y, Li J, Xi F, Li S, Li Y, Zhang Z, Yang X, Zhao M, Wang P, Guan Y, Cen Z, Zhao X, Christner M, Kobbe R, Loos S, Oh J, Yang L, Danchin A, Gao GF, Song Y, Li Y, Yang H, Wang J, Xu J, Pallen MJ, Wang J, Aepfelbacher M, Yang R, Consortium EcOHGAC-S (2011) Open-source genomic analysis of Shiga-toxin-producing *E coli* O104:H4. N Engl J Med 365(8):718–724. https://doi.org/10.1056/NEJMoa1107643

100. Tewolde R, Dallman T, Schaefer U, Sheppard CL, Ashton P, Pichon B, Ellington M, Swift C, Green J, Underwood A (2016) MOST: a modified MLST typing tool based on short read sequencing. PeerJ 4:e2308. https://doi.org/10.7717/peerj.2308

101. Walk ST, Alm EW, Calhoun LM, Mladonicky JM, Whittam TS (2007) Genetic diversity and population structure of Escherichia coli isolated from freshwater beaches.

Environ Microbiol 9(9):2274–2288. https://doi.org/10.111 1/j.1462-2920.2007.01341.x

102. Nowrouzian FL, Adlerberth I, Wold AE (2006) Enhanced persistence in the colonic microbiota of *Escherichia coli* strains belonging to phylogenetic group B2: role of virulence factors and adherence to colonic cells. Microbes Infect 8(3):834–840. https://doi.org/10.1016/j.micinf.2005.10.011

103. Clermont O, Bonacorsi S, Bingen E (2000) Rapid and simple determination of the *Escherichia coli* phylogenetic group. Appl Environ Microbiol 66(10):4555–4558. https://doi.org/10.1128/aem.66.10.4555-4558.2000

104. Beghain J, Bridier-Nahmias A, Le Nagard H, Denamur E, Clermont O (2018) ClermonTyping: an easy-to-use and accurate in silico method for Escherichia genus strain phylotyping. Microb Genom. https://doi.org/10.1099/mgen.0.000192

105. D'Costa VM, King CE, Kalan L, Morar M, Sung WW, Schwarz C, Froese D, Zazula G, Calmels F, Debruyne R, Golding GB, Poinar HN, Wright GD (2011) Antibiotic resistance is ancient. Nature 477(7365):457–461. https://doi.org/10.1038/nature10388

106. Surette MD, Wright GD (2017) Lessons from the environmental antibiotic resistome. Annu Rev Microbiol 71:309–329. https://doi.org/10.1146/annurev-micro-090816-093420

107. Sandegren L, Andersson DI (2009) Bacterial gene amplification: implications for the evolution of antibiotic resistance. Nat Rev Microbiol 7(8):578–588. https://doi.org/10.1038/nrmicro2174

108. Blair JM, Webber MA, Baylay AJ, Ogbolu DO, Piddock LJ (2015) Molecular mechanisms of antibiotic resistance. Nat Rev Microbiol 13(1):42–51. https://doi.org/10.1038/nrmicro3380

109. von Wintersdorff CJ, Penders J, van Niekerk JM, Mills ND, Majumder S, van Alphen LB, Savelkoul PH, Wolffs PF (2016) Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. Front Microbiol 7:173. https://doi.org/10.3389/fmicb.2016.00173

110. Balaban NQ, Helaine S, Lewis K, Ackermann M, Aldridge B, Andersson DI, Brynildsen MP, Bumann D, Camilli A, Collins JJ, Dehio C, Fortune S, Ghigo JM, Hardt WD, Harms A, Heinemann M, Hung DT, Jenal U, Levin BR, Michiels J, Storz G, Tan MW, Tenson T, Van Melderen L, Zinkernagel A (2019) Definitions and guidelines for research on antibiotic persistence. Nat Rev Microbiol 17(7):441–448. https://doi.org/10.1038/s41579-019-0196-3

111. Schrader SM, Vaubourgeix J, Nathan C (2020) Biology of antimicrobial resistance and approaches to combat it. Sci Transl Med. https://doi.org/10.1126/scitranslmed.aaz6992

112. Walsh C (2000) Molecular mechanisms that confer antibacterial drug resistance. Nature 406(6797):775–781. https://doi.org/10.1038/35021219

113. Boolchandani M, D'Souza AW, Dantas G (2019) Sequencing-based methods and resources to study antimicrobial resistance. Nat Rev Genet 20(6):356–370. https://doi.org/10.1038/s41576-019-0108-4

114. D'Costa VM, McGrann KM, Hughes DW, Wright GD (2006) Sampling the antibiotic resistome. Science 311(5759):374–377. https://doi.org/10.1126/science.1120800

115. Wright GD (2007) The antibiotic resistome: the nexus of chemical and genetic diversity. Nat Rev Microbiol 5(3):175–186. https://doi.org/10.1038/nrmicro1614

116. WHO (2019) WHO model lists of essential medicines. https://www.who.int/medicines/publications/essentialmedicines/en/.

117. Duke Antimicrobial Stewardship Outreach Network (2017) Developing patient safety outcome measures and measurement tools for antibiotic stewardship programs, metrics guide. https://dason.medicine.duke.edu/system/files/page-attachments/704/dason-cdcfinalanalysistool-r11-21.pdf. Accessed 15 Aug 2020

118. Yong D, Toleman MA, Giske CG, Cho HS, Sundman K, Lee K, Walsh TR (2009) Characterization of a new metallo-beta-lactamase gene, bla(NDM-1), and a novel erythromycin esterase gene carried on a unique genetic structure in Klebsiella pneumoniae sequence type 14 from India. Antimicrob Agents Chemother 53(12):5046–5054. https://doi.org/10.1128/AAC.00774-09

119. Yigit H, Queenan AM, Anderson GJ, Domenech-Sanchez A, Biddle JW, Steward CD, Alberti S, Bush K, Tenover FC (2001) Novel carbapenem-hydrolyzing beta-lactamase, KPC-1, from a carbapenem-resistant strain of Klebsiella pneumoniae. Antimicrob Agents Chemother 45(4):1151–1161. https://doi.org/10.1128/AAC.45.4.1151-1161.2001

120. Potter RF, D'Souza AW, Dantas G (2016) The rapid spread of carbapenem-resistant Enterobacteriaceae. Drug Resist Updat 29:30–46. https://doi.org/10.1016/j.drup.2016.09.002

121. Pesesky MW, Hussain T, Wallace M, Wang B, Andleeb S, Burnham CA, Dantas G (2015) KPC and NDM-1 genes in related Enterobacteriaceae strains and plasmids from Pakistan and the United States. Emerg Infect Dis 21(6):1034–1037. https://doi.org/10.3201/eid2106.141504

122. Martinez-Martinez L, Pascual A, Jacoby GA (1998) Quinolone resistance from a transferable plasmid. Lancet 351(9105):797–799. https://doi.org/10.1016/S0140-6736(97)07322-4

123. Poirel L, Rodriguez-Martinez JM, Mammeri H, Liard A, Nordmann P (2005) Origin of plasmid-mediated quinolone resistance determinant QnrA. Antimicrob Agents Chemother 49(8):3523–3525. https://doi.org/10.1128/AAC.49.8.3523-3525.2005

124. Jacoby GA, Strahilevitz J, Hooper DC (2014) Plasmid-mediated quinolone resistance. Microbiol Spectr. https://doi.org/10.1128/microbiolspec.PLAS-0006-2013

125. Gasparrini AJ, Markley JL, Kumar H, Wang B, Fang L, Irum S, Symister CT, Wallace M, Burnham CD, Andleeb S, Tolia NH, Wencewicz TA, Dantas G (2020) Tetracycline-inactivating enzymes from environmental, human commensal, and pathogenic bacteria cause broad-spectrum tetracycline resistance. Commun Biol 3(1):241. https://doi.org/10.1038/s42003-020-0966-5

126. Jiang Y, Zhang Y, Lu J, Wang Q, Cui Y, Wang Y, Zhoa D, Du X, Liu H, Li X, Wu X, Hua X, Feng Y, Yu Y (2020) Clinical relevance and plasmid dynamics of mcr-1-positive *Escherichia coli* in China: a multicentre case-control and molecular epidemiological study. Lancet Microbe 1(1):E24–E33. https://doi.org/10.1016/S2666-5247(20)30001-X

127. Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer MO, Dantas G (2012) The shared antibiotic resistome of soil bacteria and human pathogens. Science 337(6098):1107–1111. https://doi.org/10.1126/science.1220761

128. UniProt C (2015) UniProt: a hub for protein information. Nucleic Acids Res 43(Database issue):D204-212. https://doi.org/10.1093/nar/gku989

129. Crofts TS, Gasparrini AJ, Dantas G (2017) Next-generation approaches to understand and combat the antibiotic resistome. Nat Rev Microbiol 15(7):422–434. https://doi.org/10.1038/nrmicro.2017.28

130. McInnes RS, McCallum GE, Lamberte LE, van Schaik W (2020) Horizontal transfer of antibiotic resistance genes in the human gut microbiome. Curr Opin Microbiol 53:35–43. https://doi.org/10.1016/j.mib.2020.02.002

131. Sun J, Liao XP, D'Souza AW, Boolchandani M, Li SH, Cheng K, Luis Martinez J, Li L, Feng YJ, Fang LX, Huang T, Xia J, Yu Y, Zhou YF, Sun YX, Deng XB, Zeng ZL, Jiang HX, Fang BH, Tang YZ, Lian XL, Zhang RM, Fang ZW, Yan QL, Dantas G, Liu YH (2020) Environmental remodeling of human gut microbiota and antibiotic resistome in livestock farms. Nat Commun 11(1):1427. https://doi.org/10.1038/s41467-020-15222-y

132. Bhattacharyya RP, Bandyopadhyay N, Ma P, Son SS, Liu J, He LL, Wu L, Khafizov R, Boykin R, Cerqueira GC, Pironti

A, Rudy RF, Patel MM, Yang R, Skerry J, Nazarian E, Musser KA, Taylor J, Pierce VM, Earl AM, Cosimi LA, Shoresh N, Beechem J, Livny J, Hung DT (2019) Simultaneous detection of genotype and phenotype enables rapid and accurate antibiotic susceptibility determination. Nat Med 25(12):1858–1864. https://doi.org/10.1038/s41591-019-0650-9

133. van Belkum A, Welker M, Pincus D, Charrier JP, Girard V (2017) Matrix-assisted laser desorption ionization time-of-flight mass spectrometry in clinical microbiology: what are the current issues? Ann Lab Med 37(6):475–483. https://doi.org/10.3343/alm.2017.37.6.475

134. Jorgensen JH, Ferraro MJ (2009) Antimicrobial susceptibility testing: a review of general principles and contemporary practices. Clin Infect Dis 49(11):1749–1755. https://doi.org/10.1086/647952

135. Clinical and Laboratory Standards Insitute (2019) Performance standards for antimicrobial susceptibility testing, vol M100Ed29. Wayne, PA. https://clsi.org/standards/products/microbiology/documents/m100/

136. Kahlmeter G, Brown DF, Goldstein FW, MacGowan AP, Mouton JW, Osterlund A, Rodloff A, Steinbakk M, Urbaskova P, Vatopoulos A (2003) European harmonization of MIC breakpoints for antimicrobial susceptibility testing of bacteria. J Antimicrob Chemother 52(2):145–148. https://doi.org/10.1093/jac/dkg312

137. Prim N, Rivera A, Coll P, Mirelis B (2018) Is colistin susceptibility testing finally on the right track? Antimicrob Agents Chemother. https://doi.org/10.1128/AAC.02067-17

138. Liu YY, Wang Y, Walsh TR, Yi LX, Zhang R, Spencer J, Doi Y, Tian G, Dong B, Huang X, Yu LF, Gu D, Ren H, Chen X, Lv L, He D, Zhou H, Liang Z, Liu JH, Shen J (2016) 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(2):161–168. https://doi.org/10.1016/S1473-3099(15)00424-7

139. He T, Wang R, Liu D, Walsh TR, Zhang R, Lv Y, Ke Y, Ji Q, Wei R, Liu Z, Shen Y, Wang G, Sun L, Lei L, Lv Z, Li Y, Pang M, Wang L, Sun Q, Fu Y, Song H, Hao Y, Shen Z, Wang S, Chen G, Wu C, Shen J, Wang Y (2019) Emergence of plasmid-mediated high-level tigecycline resistance genes in animals and humans. Nat Microbiol 4(9):1450–1456. https://doi.org/10.1038/s41564-019-0445-2

140. Jarvinen AK, Laakso S, Piiparinen P, Aittakorpi A, Lindfors M, Huopaniemi L, Piiparinen H, Maki M (2009) Rapid identification of bacterial pathogens using a PCR- and microarray-based assay. BMC Microbiol 9:161. https://doi.org/10.1186/1471-2180-9-161

141. Poirel L, Walsh TR, Cuvillier V, Nordmann P (2011) Multiplex PCR for detection of acquired carbapenemase genes. Diagn Microbiol Infect Dis 70(1):119–123. https://doi.org/10.1016/j.diagmicrobio.2010.12.002

142. Ng LK, Martin I, Alfa M, Mulvey M (2001) Multiplex PCR for the detection of tetracycline resistant genes. Mol Cell Probes 15(4):209–215. https://doi.org/10.1006/mcpr.2001.0363

143. Liu J, Gratz J, Amour C, Kibiki G, Becker S, Janaki L, Verweij JJ, Taniuchi M, Sobuz SU, Haque R, Haverstick DM, Houpt ER (2013) A laboratory-developed TaqMan Array Card for simultaneous detection of 19 enteropathogens. J Clin Microbiol 51(2):472–480. https://doi.org/10.1128/JCM.02658-12

144. 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(11):2640–2644. https://doi.org/10.1093/jac/dks261

145. Zankari E, Allesoe 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(10):2764–2768. https://doi.org/10.1093/jac/dkx217

146. 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 FS, Wright GD, McArthur AG (2017) CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res 45(D1):D566–D573. https://doi.org/10.1093/nar/gkw1004

147. Inouye M, Dashnow H, Raven LA, 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(11):90. https://doi.org/10.1186/s13073-014-0090-6

148. Clausen PT, 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(9):2484–2488. https://doi.org/10.1093/jac/dkw184

149. Thai QK, Bos F, Pleiss J (2009) The Lactamase Engineering Database: a critical survey of TEM sequences in public databases. BMC Genomics 10:390. https://doi.org/10.1186/1471-2164-10-390

150. Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB (2009) Tuberculosis drug resistance mutation database. PLoS Med 6(2):e2. https://doi.org/10.1371/journal.pmed.1000002

151. Skov RL, Monnet DL (2016) Plasmid-mediated colistin resistance (mcr-1 gene): three months later, the story unfolds. Euro Surveill 21(9):30155. https://doi.org/10.2807/1560-7917.ES.2016.21.9.30155

152. Carnevali C, Morganti M, Scaltriti E, Bolzoni L, Pongolini S, Casadei G (2016) Occurrence of mcr-1 in colistin-resistant salmonella enterica isolates recovered from humans and animals in Italy, 2012 to 2015. Antimicrob Agents Chemother 60(12):7532–7534. https://doi.org/10.1128/AAC.01803-16

153. Ortega-Paredes D, Barba P, Zurita J (2016) Colistin-resistant Escherichia coli clinical isolate harbouring the mcr-1 gene in Ecuador. Epidemiol Infect 144(14):2967–2970. https://doi.org/10.1017/S0950268816001369

154. Teo JQ, Ong RT, Xia E, Koh TH, Khor CC, Lee SJ, Lim TP, Kwa AL (2016) mcr-1 in multidrug-resistant blaKPC-2-producing clinical enterobacteriaceae isolates in Singapore. Antimicrob Agents Chemother 60(10):6435–6437. https://doi.org/10.1128/AAC.00804-16

155. Delgado-Blas JF, Ovejero CM, Abadia-Patino L, Gonzalez-Zorn B (2016) Coexistence of mcr-1 and blaNDM-1 in *Escherichia coli* from Venezuela. Antimicrob Agents Chemother 60(10):6356–6358. https://doi.org/10.1128/AAC.01319-16

156. Wong SC, Tse H, Chen JH, Cheng VC, Ho PL, Yuen KY (2016) Colistin-resistant enterobacteriaceae carrying the mcr-1 gene among patients in Hong Kong. Emerg Infect Dis 22(9):1667–1669. https://doi.org/10.3201/eid2209.160091

157. von Wintersdorff CJ, Wolffs PF, van Niekerk JM, Beuken E, van Alphen LB, Stobberingh EE, Oude Lashof AM, Hoebe CJ, Savelkoul PH, Penders J (2016) Detection of the plasmid-mediated colistin-resistance gene mcr-1 in faecal metagenomes of Dutch travellers. J Antimicrob Chemother 71(12):3416–3419. https://doi.org/10.1093/jac/dkw328

158. Kaminski J, Gibson MK, Franzosa EA, Segata N, Dantas G, Huttenhower C (2015) High-specificity targeted functional profiling in microbial communities with ShortBRED. PLoS Comput Biol 11(12):e1004557. https://doi.org/10.1371/journal.pcbi.1004557

159. Pehrsson EC, Tsukayama P, Patel S, Mejia-Bautista M, Sosa-Soto G, Navarrete KM, Calderon M, Cabrera L, Hoyos-Arango W, Bertoli MT, Berg DE, Gilman RH, Dantas G (2016) Interconnected microbiomes and resistomes in low-income human

habitats. Nature 533(7602):212–216. https://doi.org/10.1038/nature17672

160. Gibson MK, Forsberg KJ, Dantas G (2015) Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. ISME J 9(1):207–216. https://doi.org/10.1038/ismej.2014.106

161. Eddy SR (2004) What is a hidden Markov model? Nat Biotechnol 22(10):1315–1316. https://doi.org/10.1038/nbt1004-1315

162. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD (2019) The Pfam protein families database in 2019. Nucleic Acids Res 47(D1):D427–D432. https://doi.org/10.1093/nar/gky995

163. Liu B, Pop M (2009) ARDB–antibiotic resistance genes database. Nucleic Acids Res 37(Database issue):D443-447. https://doi.org/10.1093/nar/gkn656

164. Fernandez-Arrojo L, Guazzaroni ME, Lopez-Cortes N, Beloqui A, Ferrer M (2010) Metagenomic era for biocatalyst identification. Curr Opin Biotechnol 21(6):725–733. https://doi.org/10.1016/j.copbio.2010.09.006

165. Campbell TP, Sun X, Patel VH, Sanz C, Morgan D, Dantas G (2020) The microbiome and resistome of chimpanzees, gorillas, and humans across host lifestyle and geography. ISME J 14(6):1584–1599. https://doi.org/10.1038/s41396-020-0634-2

166. Forsberg KJ, Patel S, Wencewicz TA, Dantas G (2015) The tetracycline destructases: a novel family of tetracycline-inactivating enzymes. Chem Biol 22(7):888–897. https://doi.org/10.1016/j.chembiol.2015.05.017

167. Cohen LJ, Han S, Huang YH, Brady SF (2018) Identification of the colicin V bacteriocin gene cluster by functional screening of a human microbiome metagenomic library. ACS Infect Dis 4(1):27–32. https://doi.org/10.1021/acsinfecdis.7b00081

168. Wallace JC, Port JA, Smith MN, Faustman EM (2017) FARME DB: a functional antibiotic resistance element database. Database (Oxford). https://doi.org/10.1093/database/baw165

169. Munk P, Knudsen BE, Lukjancenko O, Duarte ASR, Van Gompel L, Luiken REC, Smit LAM, Schmitt H, Garcia AD, Hansen RB, Petersen TN, Bossers A, Ruppe E, Group E, Lund O, Hald T, Pamp SJ, Vigre H, Heederik D, Wagenaar JA, Mevius D, Aarestrup FM (2018) Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries. Nat Microbiol 3(8):898-908. doi:https://doi.org/10.1038/s41564-018-0192-9

170. Barczak AK, Gomez JE, Kaufmann BB, Hinson ER, Cosimi L, Borowsky ML, Onderdonk AB, Stanley SA, Kaur D, Bryant KF, Knipe DM, Sloutsky A, Hung DT (2012) RNA signatures allow rapid identification of pathogens and antibiotic susceptibilities. Proc Natl Acad Sci USA 109(16):6217–6222. https://doi.org/10.1073/pnas.1119540109

171. Sangurdekar DP, Srienc F, Khodursky AB (2006) A classification based framework for quantitative description of large-scale microarray data. Genome Biol 7(4):R32. https://doi.org/10.1186/gb-2006-7-4-r32

172. Hrabak J, Studentova V, Walkova R, Zemlickova H, Jakubu V, Chudackova E, Gniadkowski M, Pfeifer Y, Perry JD, Wilkinson K, Bergerova T (2012) Detection of NDM-1, VIM-1, KPC, OXA-48, and OXA-162 carbapenemases by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 50(7):2441–2443. https://doi.org/10.1128/JCM.01002-12

173. Josten M, Dischinger J, Szekat C, Reif M, Al-Sabti N, Sahl HG, Parcina M, Bekeredjian-Ding I, Bierbaum G (2014) Identification of agr-positive methicillin-resistant Staphylococcus aureus harbouring the class A mec complex by MALDI-TOF

mass spectrometry. Int J Med Microbiol 304(8):1018–1023. https://doi.org/10.1016/j.ijmm.2014.07.005

174. Figueroa-Espinosa R, Costa A, Cejas D, Barrios R, Vay C, Radice M, Gutkind G, Di Conza J (2019) MALDI-TOF MS based procedure to detect KPC-2 directly from positive blood culture bottles and colonies. J Microbiol Methods 159:120–127. https://doi.org/10.1016/j.mimet.2019.02.020

175. Banyoczki G, Goebel N, Antonucci F, Zollikofer C, Stuckmann G (1990) CT diagnosis in idiopathic thrombosis of the superior mesenteric vein. Rofo 153(2):192–196. https://doi.org/10.1055/s-2008-1033360

176. Sparbier K, Lange C, Jung J, Wieser A, Schubert S, Kostrzewa M (2013) MALDI biotyper-based rapid resistance detection by stable-isotope labeling. J Clin Microbiol 51(11):3741–3748. https://doi.org/10.1128/JCM.01536-13

177. Jung JS, Hamacher C, Gross B, Sparbier K, Lange C, Kostrzewa M, Schubert S (2016) Evaluation of a semiquantitative matrix-assisted laser desorption ionization-time of flight mass spectrometry method for rapid antimicrobial susceptibility testing of positive blood cultures. J Clin Microbiol 54(11):2820–2824. https://doi.org/10.1128/JCM.01131-16

178. Cawthon RM, Weiss R, Xu GF, Viskochil D, Culver M, Stevens J, Robertson M, Dunn D, Gesteland R, O'Connell P et al (1990) A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations. Cell 62(1):193–201. https://doi.org/10.1016/0092-8674(90)90253-b

179. Hatosy SM, Martiny AC (2015) The ocean as a global reservoir of antibiotic resistance genes. Appl Environ Microbiol 81(21):7593–7599. https://doi.org/10.1128/AEM.00736-15

180. Rizzo L, Manaia C, Merlin C, Schwartz T, Dagot C, Ploy MC, Michael I, Fatta-Kassinos D (2013) Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: a review. Sci Total Environ 447:345–360. https://doi.org/10.1016/j.scitotenv.2013.01.032

181. Evans DR, Griffith MP, Sundermann AJ, Shutt KA, Saul MI, Mustapha MM, Marsh JW, Cooper VS, Harrison LH, Van Tyne D (2020) Systematic detection of horizontal gene transfer across genera among multidrug-resistant bacteria in a single hospital. Elife. https://doi.org/10.7554/eLife.53886

182. Lerat E, Ochman H (2004) Psi-Phi: exploring the outer limits of bacterial pseudogenes. Genome Res 14(11):2273–2278. https://doi.org/10.1101/gr.2925604

183. Carr VR, Shkoporov A, Hill C, Mullany P, Moyes DL (2020) Probing the mobilome: discoveries in the dynamic microbiome. Trends Microbiol. https://doi.org/10.1016/j.tim.2020.05.003

184. Jorgensen TS, Kiil AS, Hansen MA, Sorensen SJ, Hansen LH (2014) Current strategies for mobilome research. Front Microbiol 5:750. https://doi.org/10.3389/fmicb.2014.00750

185. Carattoli A (2013) Plasmids and the spread of resistance. Int J Med Microbiol 303(6–7):298–304. https://doi.org/10.1016/j.ijmm.2013.02.001

186. Smalla K, Jechalke S, Top EM (2015) Plasmid detection, characterization, and ecology. Microbiol Spectr 3 (1):PLAS-0038–2014. doi:https://doi.org/10.1128/microbiolspec.PLAS-0038-2014

187. Pinilla-Redondo R, Cyriaque V, Jacquiod S, Sorensen SJ, Riber L (2018) Monitoring plasmid-mediated horizontal gene transfer in microbiomes: recent advances and future perspectives. Plasmid 99:56–67. https://doi.org/10.1016/j.plasmid.2018.08.002

188. Emamalipour M, Seidi K, Zununi Vahed S, Jahanban-Esfahlan A, Jaymand M, Majdi H, Amoozgar Z, Chitkushev LT, Javaheri T, Jahanban-Esfahlan R, Zare P (2020) Horizontal gene transfer: from evolutionary flexibility to disease progression. Front Cell Dev Biol 8:229. https://doi.org/10.3389/fcell.2020.00229

189. Conlan S, Park M, Deming C, Thomas PJ, Young AC, Coleman H, Sison C, Program NCS, Weingarten RA, Lau AF, Dekker JP,

Palmore TN, Frank KM, Segre JA (2016) Plasmid Dynamics in KPC-Positive Klebsiella pneumoniae during Long-Term Patient Colonization. mBio. https://doi.org/10.1128/mBio.00742-16

190. Partridge SR, Kwong SM, Firth N, Jensen SO (2018) Mobile genetic elements associated with antimicrobial resistance. Clin Microbiol Rev. https://doi.org/10.1128/CMR.00088-17

191. Carattoli A, Bertini A, Villa L, Falbo V, Hopkins KL, Threlfall EJ (2005) Identification of plasmids by PCR-based replicon typing. J Microbiol Methods 63(3):219–228. https://doi.org/10.1016/j.mimet.2005.03.018

192. Garcia-Fernandez A, Villa L, Moodley A, Hasman H, Miriagou V, Guardabassi L, Carattoli A (2011) Multilocus sequence typing of IncN plasmids. J Antimicrob Chemother 66(9):1987–1991. https://doi.org/10.1093/jac/dkr225

193. Valverde A, Canton R, Garcillan-Barcia MP, Novais A, Galan JC, Alvarado A, de la Cruz F, Baquero F, Coque TM (2009) Spread of bla(CTX-M-14) is driven mainly by IncK plasmids disseminated among Escherichia coli phylogroups A, B1, and D in Spain. Antimicrob Agents Chemother 53(12):5204–5212. https://doi.org/10.1128/AAC.01706-08

194. Alvarado A, Garcillan-Barcia MP, de la Cruz F (2012) A degenerate primer MOB typing (DPMT) method to classify gamma-proteobacterial plasmids in clinical and environmental settings. PLoS ONE 7(7):e40438. https://doi.org/10.1371/journal.pone.0040438

195. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, Moller Aarestrup F, Hasman H (2014) In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother 58(7):3895–3903. https://doi.org/10.1128/AAC.02412-14

196. Francia MV, Varsaki A, Garcillan-Barcia MP, Latorre A, Drainas C, de la Cruz F (2004) A classification scheme for mobilization regions of bacterial plasmids. FEMS Microbiol Rev 28(1):79–100. https://doi.org/10.1016/j.femsre.2003.09.001

197. Hazen TH, Mettus R, McElheny CL, Bowler SL, Nagaraj S, Doi Y, Rasko DA (2018) Diversity among blaKPC-containing plasmids in Escherichia coli and other bacterial species isolated from the same patients. Sci Rep 8(1):10291. https://doi.org/10.1038/s41598-018-28085-7

198. Pecora ND, Li N, Allard M, Li C, Albano E, Delaney M, Dubois A, Onderdonk AB, Bry L (2015) Genomically informed surveillance for carbapenem-resistant enterobacteriaceae in a health care system. mBio 6(4):e01030. https://doi.org/10.1128/mBio.01030-15

199. Orlek A, Phan H, Sheppard AE, Doumith M, Ellington M, Peto T, Crook D, Walker AS, Woodford N, Anjum MF, Stoesser N (2017) Ordering the mob: Insights into replicon and MOB typing schemes from analysis of a curated dataset of publicly available plasmids. Plasmid 91:42–52. https://doi.org/10.1016/j.plasmid.2017.03.002

200. Arredondo-Alonso S, Willems RJ, van Schaik W, Schurch AC (2017) On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. Microb Genom 3(10):e000128. https://doi.org/10.1099/mgen.0.000128

201. Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, Pevzner PA (2016) plasmidSPAdes: assembling plasmids from whole genome sequencing data. Bioinformatics 32(22):3380–3387. https://doi.org/10.1093/bioinformatics/btw493

202. Zhou F, Xu Y (2010) cBar: a computer program to distinguish plasmid-derived from chromosome-derived sequence fragments in metagenomics data. Bioinformatics 26(16):2051–2052. https://doi.org/10.1093/bioinformatics/btq299

203. Krawczyk PS, Lipinski L, Dziembowski A (2018) PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. Nucleic Acids Res 46(6):e35. https://doi.org/10.1093/nar/gkx1321

204. Rozov R, Brown Kav A, Bogumil D, Shterzer N, Halperin E, Mizrahi I, Shamir R (2017) Recycler: an algorithm for detecting plasmids from de novo assembly graphs. Bioinformatics 33(4):475–482. https://doi.org/10.1093/bioinformatics/btw651

205. Antipov D, Raiko M, Lapidus A, Pevzner PA (2019) Plasmid detection and assembly in genomic and metagenomic data sets. Genome Res 29(6):961–968. https://doi.org/10.1101/gr.241299.118

206. Orlek A, Stoesser N, Anjum MF, Doumith M, Ellington MJ, Peto T, Crook D, Woodford N, Walker AS, Phan H, Sheppard AE (2017) Plasmid classification in an Era of whole-genome sequencing: application in studies of antibiotic resistance epidemiology. Front Microbiol 8:182. https://doi.org/10.3389/fmicb.2017.00182

207. Sheppard AE, Stoesser N, Wilson DJ, Sebra R, Kasarskis A, Anson LW, Giess A, Pankhurst LJ, Vaughan A, Grim CJ, Cox HL, Yeh AJ, Modernising Medical Microbiology Informatics G, Sifri CD, Walker AS, Peto TE, Crook DW, Mathers AJ (2016) Nested Russian doll-like genetic mobility drives rapid dissemination of the carbapenem resistance gene blaKPC. Antimicrob Agents Chemother 60 (6):3767-3778. doi:https://doi.org/10.1128/AAC.00464-16

208. Galata V, Fehlmann T, Backes C, Keller A (2019) PLSDB: a resource of complete bacterial plasmids. Nucleic Acids Res 47(D1):D195–D202. https://doi.org/10.1093/nar/gky1050

209. Stoesser N, Giess A, Batty EM, Sheppard AE, Walker AS, Wilson DJ, Didelot X, Bashir A, Sebra R, Kasarskis A, Sthapit B, Shakya M, Kelly D, Pollard AJ, Peto TE, Crook DW, Donnelly P, Thorson S, Amatya P, Joshi S (2014) Genome sequencing of an extended series of NDM-producing Klebsiella pneumoniae isolates from neonatal infections in a Nepali hospital characterizes the extent of community- versus hospital-associated transmission in an endemic setting. Antimicrob Agents Chemother 58(12):7347–7357. https://doi.org/10.1128/AAC.03900-14

210. Conlan S, Thomas PJ, Deming C, Park M, Lau AF, Dekker JP, Snitkin ES, Clark TA, Luong K, Song Y, Tsai YC, Boitano M, Dayal J, Brooks SY, Schmidt B, Young AC, Thomas JW, Bouffard GG, Blakesley RW, Program NCS, Mullikin JC, Korlach J, Henderson DK, Frank KM, Palmore TN, Segre JA (2014) Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. Sci Transl Med 6(254):254. https://doi.org/10.1126/scitranslmed.3009845

211. Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW (2010) A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res 38(15):e159. https://doi.org/10.1093/nar/gkq543

212. Cazares A, Moore MP, Hall JPJ, Wright LL, Grimes M, Emond-Rheault JG, Pongchaikul P, Santanirand P, Levesque RC, Fothergill JL, Winstanley C (2020) A megaplasmid family driving dissemination of multidrug resistance in Pseudomonas. Nat Commun 11(1):1370. https://doi.org/10.1038/s41467-020-15081-7

213. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M (2006) ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34(Database issue):D32-36. https://doi.org/10.1093/nar/gkj014

214. Hawkey J, Hamidian M, Wick RR, Edwards DJ, Billman-Jacobe H, Hall RM, Holt KE (2015) ISMapper: identifying transposase insertion sites in bacterial genomes from short read sequence data. BMC Genomics 16:667. https://doi.org/10.1186/s12864-015-1860-2

215. Cury J, Jove T, Touchon M, Neron B, Rocha EP (2016) Identification and analysis of integrons and cassette arrays in bacterial genomes. Nucleic Acids Res 44(10):4539–4550. https://doi.org/10.1093/nar/gkw319

216. Treepong P, Guyeux C, Meunier A, Couchoud C, Hocquet D, Valot B (2018) panISa: ab initio detection of insertion sequences in bacterial genomes from short read sequence data. Bioinformatics 34(22):3795–3800. https://doi.org/10.1093/bioinformatics/bty479

217. Durrant MG, Li MM, Siranosian BA, Montgomery SB, Bhatt AS (2020) A bioinformatic analysis of integrative mobile genetic elements highlights their role in bacterial adaptation. Cell Host Microbe 27(1):140-153 e149. https://doi.org/10.1016/j.chom.2019.10.022

218. Jorgensen TS, Xu Z, Hansen MA, Sorensen SJ, Hansen LH (2014) Hundreds of circular novel plasmids and DNA elements identified in a rat cecum metamobilome. PLoS ONE 9(2):e87924. https://doi.org/10.1371/journal.pone.0087924

219. Brown Kav A, Benhar I, Mizrahi I (2013) A method for purifying high quality and high yield plasmid DNA for metagenomic and deep sequencing approaches. J Microbiol Methods 95(2):272–279. https://doi.org/10.1016/j.mimet.2013.09.008

220. Li LL, Norman A, Hansen LH, Sorensen SJ (2012) Metamobilomics–expanding our knowledge on the pool of plasmid encoded traits in natural environments using high-throughput sequencing. Clin Microbiol Infect 18(Suppl 4):5–7. https://doi.org/10.1111/j.1469-0691.2012.03862.x

221. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326(5950):289–293. https://doi.org/10.1126/science.1181369

222. DeMaere MZ, Darling AE (2019) bin3C: exploiting Hi-C sequencing data to accurately resolve metagenome-assembled genomes. Genome Biol 20(1):46. https://doi.org/10.1186/s13059-019-1643-1

223. Press MO, Wiser AH, Kronenberg ZN, Langford KW, Shakya M, Lo C, Mueller KA, Sullivan S, Chain PSG, Liachko I (2017) Hi-C deconvolution of a human gut microbiome yields high-quality draft genomes and reveals plasmid-genome interactions. bioRxiv. https://doi.org/10.1101/198713

224. DeMaere MZ, Liu MYZ, Lin E, Djordjevic SP, Charles IG, Worden P, Burke CM, Monahan LG, Gardiner M, Borody TJ, Darling AE (2020) Metagenomic Hi-C of a healthy human fecal microbiome transplant donor. Microbiol Resour Announc. https://doi.org/10.1128/MRA.01523-19

225. Beitel CW, Froenicke L, Lang JM, Korf IF, Michelmore RW, Eisen JA, Darling AE (2014) Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. PeerJ 2:e415. https://doi.org/10.7717/peerj.415

226. Burton JN, Liachko I, Dunham MJ, Shendure J (2014) Specieslevel deconvolution of metagenome assemblies with Hi-C-based contact probability maps. G3 (Bethesda) 4(7):1339–1346. https://doi.org/10.1534/g3.114.011825

227. Marbouty M, Cournac A, Flot JF, Marie-Nelly H, Mozziconacci J, Koszul R (2014) Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. Elife 3:e03318. https://doi.org/10.7554/eLife.03318

228. Stewart RD, Auffret MD, Warr A, Wiser AH, Press MO, Langford KW, Liachko I, Snelling TJ, Dewhurst RJ, Walker AW, Roehe R, Watson M (2018) Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. Nat Commun 9(1):870. https://doi.org/10.1038/s41467-018-03317-6

229. Yaffe E, Relman DA (2020) Tracking microbial evolution in the human gut using Hi-C reveals extensive horizontal gene transfer, persistence and adaptation. Nat Microbiol 5(2):343–353. https://doi.org/10.1038/s41564-019-0625-0

230. Stalder T, Press MO, Sullivan S, Liachko I, Top EM (2019) Linking the resistome and plasmidome to the microbiome. ISME J 13(10):2437–2446. https://doi.org/10.1038/s41396-019-0446-4

231. Kent AG, Vill AC, Shi Q, Satlin MJ, Brito IL (2020) Widespread transfer of mobile antibiotic resistance genes within individual gut microbiomes revealed through bacterial Hi-C. Nat Commun 11(1):4379. https://doi.org/10.1038/s41467-020-18164-7

232. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, Group NCSP, Henderson DK, Palmore TN, Segre JA (2012) Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. Sci Transl Med 4(148):148ra116. Doi:https://doi.org/10.1126/scitranslmed.3004129

233. Smibert OC, Aung AK, Woolnough E, Carter GP, Schultz MB, Howden BP, Seemann T, Spelman D, McGloughlin S, Peleg AY (2018) Mobile phones and computer keyboards: unlikely reservoirs of multidrug-resistant organisms in the tertiary intensive care unit. J Hosp Infect 99(3):295–298. https://doi.org/10.1016/j.jhin.2018.02.013

234. Guerra-Assuncao JA, Houben RM, Crampin AC, Mzembe T, Mallard K, Coll F, Khan P, Banda L, Chiwaya A, Pereira RP, McNerney R, Harris D, Parkhill J, Clark TG, Glynn JR (2015) Recurrence due to relapse or reinfection with Mycobacterium tuberculosis: a whole-genome sequencing approach in a large, population-based cohort with a high HIV infection prevalence and active follow-up. J Infect Dis 211(7):1154–1163. https://doi.org/10.1093/infdis/jiu574

235. Donkor ES, Jamrozy D, Mills RO, Dankwah T, Amoo PK, Egyir B, Badoe EV, Twasam J, Bentley SD (2018) A genomic infection control study for Staphylococcus aureus in two Ghanaian hospitals. Infect Drug Resist 11:1757–1765. https://doi.org/10.2147/IDR.S167639

236. Sharma H, Ong MR, Ready D, Coelho J, Groves N, Chalker V, Warren S (2019) Real-time whole genome sequencing to control a Streptococcus pyogenes outbreak at a national orthopaedic hospital. J Hosp Infect 103(1):21–26. https://doi.org/10.1016/j.jhin.2019.07.003

237. Sundermann AJ, Babiker A, Marsh JW, Shutt KA, Mustapha MM, Pasculle AW, Ezeonwuka C, Saul MI, Pacey MP, Van Tyne D, Ayres AM, Cooper VS, Snyder GM, Harrison LH (2020) Outbreak of vancomycin-resistant enterococcus faecium in interventional radiology: detection through whole-genome sequencing-based surveillance. Clin Infect Dis 70(11):2336–2343. https://doi.org/10.1093/cid/ciz666