

Influence of Sequencing Technology on Pangenome-Level Analysis and Detection of Antimicrobial Resistance Genes in ESKAPE Pathogens

Alba Frias-De-Diego,[●] Manuel Jara,[●] and Cristina Lanzas[●]

Department of Population Health and Pathobiology, College of Veterinary Medicine, North Carolina State University, Raleigh, North Carolina, USA

As sequencing costs decrease, short-read and long-read technologies are indispensable tools for uncovering the genetic drivers behind bacterial pathogen resistance. This study explores the differences between the use of short-read (Illumina) and long-read (Oxford Nanopore Technologies [ONT]) sequencing in detecting antimicrobial resistance (AMR) genes in ESKAPE pathogens (ie, *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter cloacae*). Utilizing a dataset of 1385 whole genome sequences and applying commonly used bioinformatic methods in bacterial genomics, we assessed the differences in genomic completeness, pangenome structure, and AMR gene and point mutation identification. Illumina presented higher genome completeness, while ONT identified a broader pangenome. Hybrid assembly outperformed both Illumina and ONT at identifying key AMR genetic determinants, presented results closer to Illumina's completeness, and revealed ONT-like pangenomic content. Notably, Illumina consistently detected more AMR-related point mutations than its counterparts. This highlights the importance of method selection based on research goals, particularly when using publicly available data ranging a wide timespan. Differences were also observed for specific gene classes and bacterial species, underscoring the need for a nuanced understanding of technology limitations. Overall, this study reveals the strengths and limitations of each approach, advocating for the use of Illumina for common AMR analysis, ONT for studying complex genomes and novel species, and hybrid assembly for a more comprehensive characterization, leveraging the benefits of both technologies.

Keywords. antimicrobial resistance; bacterial genomics; ESKAPE pathogens; Illumina; Oxford Nanopore Technologies.

An estimated 4.95 million annual deaths are caused by antimicrobial-resistant pathogens worldwide [1]. These deaths are commonly caused by a group of pathogens identified as ESKAPE for their ability to resist antibiotics and “escape” treatment, creating an ongoing problem for modern medicine [2]. This group of bacteria includes *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species [2, 3]. ESKAPE pathogens are known to cause healthcare-associated infections [4]. The compromised health

statuses of patients in healthcare settings make them particularly vulnerable to infections caused by these pathogens [5]. However, ESKAPE pathogens can also be transmitted within community settings, making them widely distributed throughout populations [6].

Next-generation sequencing technologies have emerged as a critical component of clinical microbiology and antimicrobial resistance (AMR) surveillance. Whole genome sequencing of clinical isolates is extensively used for pathogen typing, resistance gene identification, and evolutionary and disease transmission studies [7–12]. At the heart of this technological revolution lie 2 types of sequencing technology with distinct strengths and weaknesses: short-read and long-read sequencing. Short-read sequencing, typified by platforms like Illumina, furnishes rapid and cost-effective data acquisition [13, 14]. This technology achieves high-depth coverage, or the average number of times each nucleotide or region of interest is sequenced in a genome, facilitating the identification of prevalent genetic variants and mutations within populations. However, its limited read length can pose challenges in accurately reconstructing complex genomic regions, such as repetitive elements, which are integral to understanding pathogen evolution [13, 15]. In contrast, long-read sequencing,

Received 04 March 2025; editorial decision 20 March 2025; accepted 24 March 2025; published online 26 March 2025

Correspondence: Alba Frias-De-Diego, PhD, Department of Population Health and Pathobiology, College of Veterinary Medicine, North Carolina State University, 1060 William Moore Dr., Raleigh, NC 27606 (afriasd@ncsu.edu).

Open Forum Infectious Diseases[®]

© The Author(s) 2025. Published by Oxford University Press on behalf of Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.
<https://doi.org/10.1093/ofid/ofaf183>

epitomized by rapidly growing platforms such as Oxford Nanopore Technologies (ONT), generates extensive reads that span challenging genomic segments [16, 17]. This characteristic empowers the accurate delineation of repetitive regions, mobile genetic elements, and structural variations that are instrumental in shaping the genomic landscape of pathogens. Moreover, long-read sequencing offers the potential to uncover novel genetic elements and achieve more comprehensive genome assembly [17–20]. However, long-read sequencing exhibits a comparatively higher error rate in raw sequences than standard next-generation sequencers like Illumina [21–23]. This discrepancy stems from ONT’s challenges in distinguishing between purines (A and G) and pyrimidines (C and T), leading to increased substitution errors, particularly in comparison to transversions [21, 24]. Additionally, low-complexity regions, homopolymers, and short tandem repeats pose further challenges for ONT, potentially inflating error rates. With a mean global error rate on raw reads estimated at approximately 6%, the accuracy of ONT heavily relies on the performance of its base-calling software [21].

An additional impacting factor on both Illumina and ONT assemblies and gene calling is based on the differences in read lengths. In the case of Illumina, shorter read lengths lead to more fragmented assemblies and, thus, less accurate gene calls. Similarly, there are differences between the library preparation and sequencing kits used for ONT, where ligation presents longer fragments than the rapid kits at the tradeoff of missing small plasmids, while newer kits have increased accuracy as compared to older materials [25–27].

The potential differences on the use of both Illumina and ONT are already being considered for research on these bacteria, with Illumina more commonly used for molecular characterization, metagenomics, and epidemiological surveillance [28–30], and ONT mostly used for real-time detection, assembly of bacterial whole genomes, and rapid diagnosis [31, 32]. However, oftentimes the decision of whether to use short- or long-read sequencing is hard to determine since only few articles have quantitatively compared these differences [33–36].

Despite the higher error rates of long-read sequencing, the combination of both long and short reads has proven efficient in overcoming this problem and improving the accuracy of the obtained results in a novel method known as hybrid assembly correction, which is increasingly being used for genomic studies [37, 38]. However, as technology advances, specifically designed bioinformatic pipelines for handling both short and long reads will gradually bridge the gap between the 2 methods. This study seeks to uncover potential disparities between short-read (Illumina) and long-read (ONT) technologies in detecting and identifying AMR genes on ESKAPE pathogens. In addition, we evaluated the ability of both technologies to describe the pangenome, the complete set of genes within each species composed of both a central or core genome (general for all

strains) and an accessory genome (particular of specific strains or even unique) [39–41]. The accessory genome is known to have a strong influence on phenotypic diversity, including genes related to drug resistance [42, 43]. For that reason, pan-genome analyses are a powerful approach to identify genetic and functional diversity of bacterial populations [39–41, 44, 45]. The results of this study aim to provide quantitative information that could help researchers decide which approach would be more suitable for their specific research. Here, we collected whole genome sequences of 1385 isolates from ESKAPE bacteria sequenced by both short- and long-read technologies and performed a series of quality control (QC) and commonly used genomic analyses to assess the potential differences associated with each specific method.

METHODS

Data Collection and Curation

In this study, we collated a comprehensive dataset composed of publicly available whole genome sequences pertaining to ESKAPE pathogens (*E faecium* [genome size between 2.5 and 3.2 Mb], *S aureus* [genome size around 2.8 Mb], *K pneumoniae* [genome size between 5.3 and 5.7 Mb], *A baumannii* [genome size between 3.6 and 4.2 Mb], *P aeruginosa* [genome size about 6.3 Mb], and *Enterobacter* species [genome size between 4.5 and 5.52 Mb]), generated by the collection of sequences from BioProjects that contained sequencing data for both Illumina and ONT. This dataset, comprising a total of 1385 sequences (after QC; [Supplementary Table 1](#)) per sequencer, was acquired from the National Center for Biotechnology Information (NCBI) Sequence Read Archive repository [46]. Isolates were only used for further analyses if they were sequenced by both Illumina and ONT sequencers. Metadata associated with each isolate including NCBI biosample, bioproject ID, sequencer, and sequencing depth are available in [Supplementary Tables 1 and 2](#).

Genome Completeness Analysis

Genome completeness was assessed using BUSCO (benchmarking universal single-copy orthologs) v4.0 [47] and comparing the selected sequences after the QC analysis and against the bacteria_odb10 database.

Bactopia Pipeline

A graphical visualization of the complete pipeline is found in [Figure 1](#).

To prepare the raw sequencing data for downstream analysis, we employed the fastq-dl tool v2..4 (available at: <https://github.com/rpetit3/fastq-dl>). The raw reads were processed using the Bactopia v3 pipeline [48], which uses different tools depending on whether the sequences are short reads (Illumina) or long reads (ONT). To ensure high-quality sequencing data, we

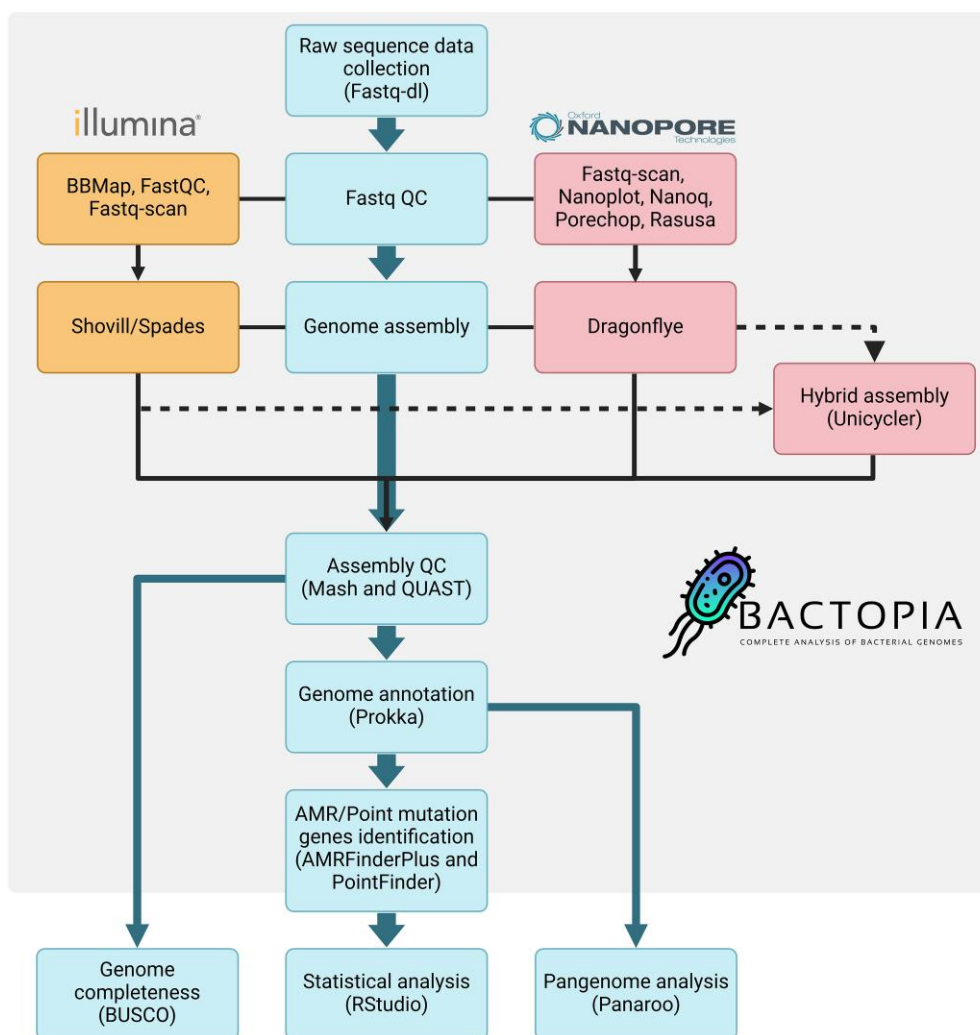


Figure 1. Complete workflow for performing antimicrobial resistance gene identification in ESKAPE pathogens using a pangenome approach, as well as statistical analysis to identify the potential differences associated with the sequencing technology (ie, Illumina and Oxford Nanopore Technologies) and hybrid method. Abbreviations: AMR, antimicrobial resistance; BUSCO, benchmarking universal single-copy orthologs; QC, quality control.

employed a comprehensive QC pipeline for Illumina and ONT samples. For Illumina sequences, FastQC v0.12.1 [49] provided detailed quality metrics, while BBMap v39.03 [50] filtered and trimmed reads with a quality score below Q20. Fastq-scan v1.0.1 (available at: <https://github.com/rpetit3/fastq-scan>) was used for a quick quality overview. For the ONT isolates, Fastq-scan [51] provided initial quality summaries, NanoPlot v1.42.0 [52] generated detailed quality plots, and Nanoq v0.10.0 [53] filtered low-quality reads. Porechop v0.2.4 [54] trimmed adapters and split chimeric reads, and Rasusa v0.7.1 [55] subsampled reads to a specified coverage. All high-quality reads were then assembled, in the case of Illumina samples using Shovill v1.1.0 [56], an integral component of the Bactopia pipeline, which leverages the SPAdes v3.15.5 assembler [57], and using Dragonflye v1.1.2 (available at: <https://github.com/rpetit3/dragonflye>) for the ONT samples. To evaluate the

assembly quality, assessments were conducted employing Mash v2.3.0 [58] and QUASt v5.2.0 [59]. The genome annotation process was executed using Prokka v1.14.0 [60]. To leverage the strengths of both short-read and long-read technologies while minimizing their limitations, we performed a hybrid assembly (short-read polishing) using Unicycler v0.5.0 [61] in Bactopia. All of the annotated sequences (.gff files) were analyzed for AMR gene content using AMRFinderPlus [62] as part of the Bactopia pipeline using both nucleotide (BLASTX) and protein (BLASTP) strategies; all matches with identity and coverage below 90% were excluded for further analysis. In addition, point mutations associated with AMR were identified using PointFinder [63].

The presence of AMR genes was compared across sequencing methods (Illumina, ONT, hybrid) using Pearson χ^2 tests with Yates' continuity correction implemented in R software

[64]. In addition, the relationship between the presence of AMR genes and point mutations, and the log-transformed total number of bases sequenced per isolate, was assessed through linear regression analysis in R software (v 4.4.3) [64].

Pangenome Analysis

To provide an additional level of genomic comparison between the Illumina and ONT isolates, we reconstructed the pangenome of each ESKAPE species using Panaroo (v1.2.2) with the strict clean-mode setting [65]. To ensure a robust and accurate gene clustering, we set the sequence identity threshold to 95%, meaning that genes needed to have at least 95% sequence similarity to be considered the same gene. This approach allowed us to identify genes that are highly conserved within the population (core genes) as well as those that are more variable (soft-core, shell, and cloud genes). Our threshold for the core genome was set at (>99% of gene similarity among all sequences) to ensure the inclusion of unique gene alleles in the accessory genome. Additionally, we identified the rest of gene types with the following thresholds: soft-core genes (95%–99%), shell genes (15%–95%), and cloud genes (<15%).

To further investigate the differences in pangenome size and composition, and assess whether the variations observed were due to differences in annotation, we performed 1-way analysis of variance (ANOVA) to compare the number of annotated genes prior to pangenome reconstruction using R software (v 4.4.3) [64].

To compare the differences in full-length versus partial genome hits across the 6 ESKAPE species, we analyzed the results generated from AMRFinderPlus. For each species, the proportions of full-length and partial genome hits were calculated for each sequencing method. We employed Pearson χ^2 test with Yates’ continuity correction to assess the statistical significance of differences in these proportions. In addition, pairwise comparisons were made between each combination of sequencing methods (Illumina vs ONT, Illumina vs hybrid, and ONT vs hybrid).

Finally, to assess the association between sequencing depth (log-transformed total number of bases sequenced) and the number of identified AMR genes and point mutations, a linear regression was performed. This analysis aimed to evaluate the relationship between sequencing deepness and the number of detected genes across all samples. Subsequently, ANOVA was conducted to compare the number of AMR genes and point mutations detected across different sequencing platforms. Sequencing models were categorized into 2 groups: ONT, including MinION, GridION, and PromethION; and Illumina, including MiniSeq, MiSeq, NextSeq, NovaSeq, iSeq, and HiSeq. The number of detected AMR genes was the dependent variable, with sequencer model as the main factor and platform (Illumina and ONT) as a covariate. An interaction term between sequencer and platform was also included to evaluate

whether the effect of sequencing models on gene detection varied based on the sequencing platform. All analyses were performed in R software (v 4.4.3) [64].

RESULTS

Overall, all the identified 1385 sequences were used as no sequences were flagged as poor quality. From those sequences, 91 isolates belonged to *E faecium*, 201 to *S aureus*, 537 to *K pneumoniae*, 257 to *A baumannii*, 159 to *P aeruginosa*, and 140 to *Enterobacter cloacae*.

Genome Completeness

The assessment of genome completeness using BUSCO showed that isolates sequenced by Illumina can reach a higher level of completeness than those sequenced by ONT. The application of hybrid assembly reached similar completeness levels than those observed for Illumina (Table 1, Supplementary Table 2).

Overall, the results of the comparison of the differences in full-length versus partial genome hits show that Illumina sequencing provides reliable full-length hits with low percentage of partial hits. Even though ONT sequences presented high full-length genome coverage, they also showed a high proportion of partial hits. In contrast, the hybrid method offers a balanced approach, combining the strengths of both Illumina and ONT to

Table 1. Genome Completeness (Based on BUSCO Results) Determined on Illumina, Oxford Nanopore Technologies, and Hybrid Method on ESKAPE Pathogens

Species	Sequencer	No.	C (>90%) ^a	C (%) ^b
<i>Enterococcus faecium</i>	Illumina	91	91	100.0
	ONT	91	21	23.1
	Hybrid	91	91	100.0
<i>Staphylococcus aureus</i>	Illumina	201	201	100.0
	ONT	201	170	84.6
	Hybrid	201	201	100.0
<i>Klebsiella pneumoniae</i>	Illumina	537	515	95.9
	ONT	537	16	3.0
	Hybrid	537	501	93.3
<i>Acinetobacter baumannii</i>	Illumina	257	256	99.6
	ONT	257	89	34.6
	Hybrid	257	257	100.0
<i>Pseudomonas aeruginosa</i>	Illumina	159	169	99.4
	ONT	159	86	54.1
	Hybrid	159	168	99.4
<i>Enterobacter cloacae</i>	Illumina	140	135	96.4
	ONT	140	13	9.3
	Hybrid	140	132	94.3

Abbreviation: BUSCO, benchmarking universal single-copy orthologs; ONT, Oxford Nanopore Technologies.

^aC (>90%) corresponds to the number of sequences surpassing the 90% threshold of coverage as indicated by the completeness metric from BUSCO.

^bC (%) indicates the percentage of the total number of sequences that surpassed that 90% threshold.

achieve comprehensive and accurate identification of AMR genes (Supplementary Figure 1).

AMR Gene Profiles Across ESKAPE Pathogens

Overall, we observed a wide diversity of resistance profiles among ESKAPE pathogens. All 3 methods (short reads, long reads, and hybrid assembly) detected a wide variety of resistance genes (detailed list available in Supplementary Table 2). The general trend across the 6 species consisted of hybrid assembly detecting an overall higher number of AMR genes, followed by Illumina and ONT (Supplementary Table 2).

Overall AMR Gene and Point Mutation Detection. We observed statistically significant differences ($P < .05$) in AMR detection between the 3 methods (panel A in Figures 2–7; Supplementary Table 3). For all species, hybrid assembly showed the highest detection power for AMR genes, outperforming both Illumina and ONT. The detection power of each technology is calculated by comparing the percentage of

times each AMR gene or point mutation was detected by each technology (Supplementary Tables 2 and 3). β -lactam resistance emerged as the AMR class with the highest number of genes identified by all 3 sequencing methods following the same pattern where hybrid assembly was the technology detecting a higher number of genes, followed by Illumina and ONT, except for *E. faecium* (0 genes detected) and *S. aureus* (7 genes), where Illumina detected more genes than both hybrid assembly and ONT.

When the overall detection of point mutations was compared, we observed that Illumina exhibited higher detection power compared to both hybrid assembly and ONT except for *S. aureus*, where hybrid assembly detected a higher number. Although Illumina was generally able to detect a higher number of point mutations, hybrid correction performed comparably well in most species with minimal differences found in its detection rates (Supplementary Tables 2 and 3). When we combined the detected AMR genes with the AMR-associated point mutations and compared them with the total number of bases

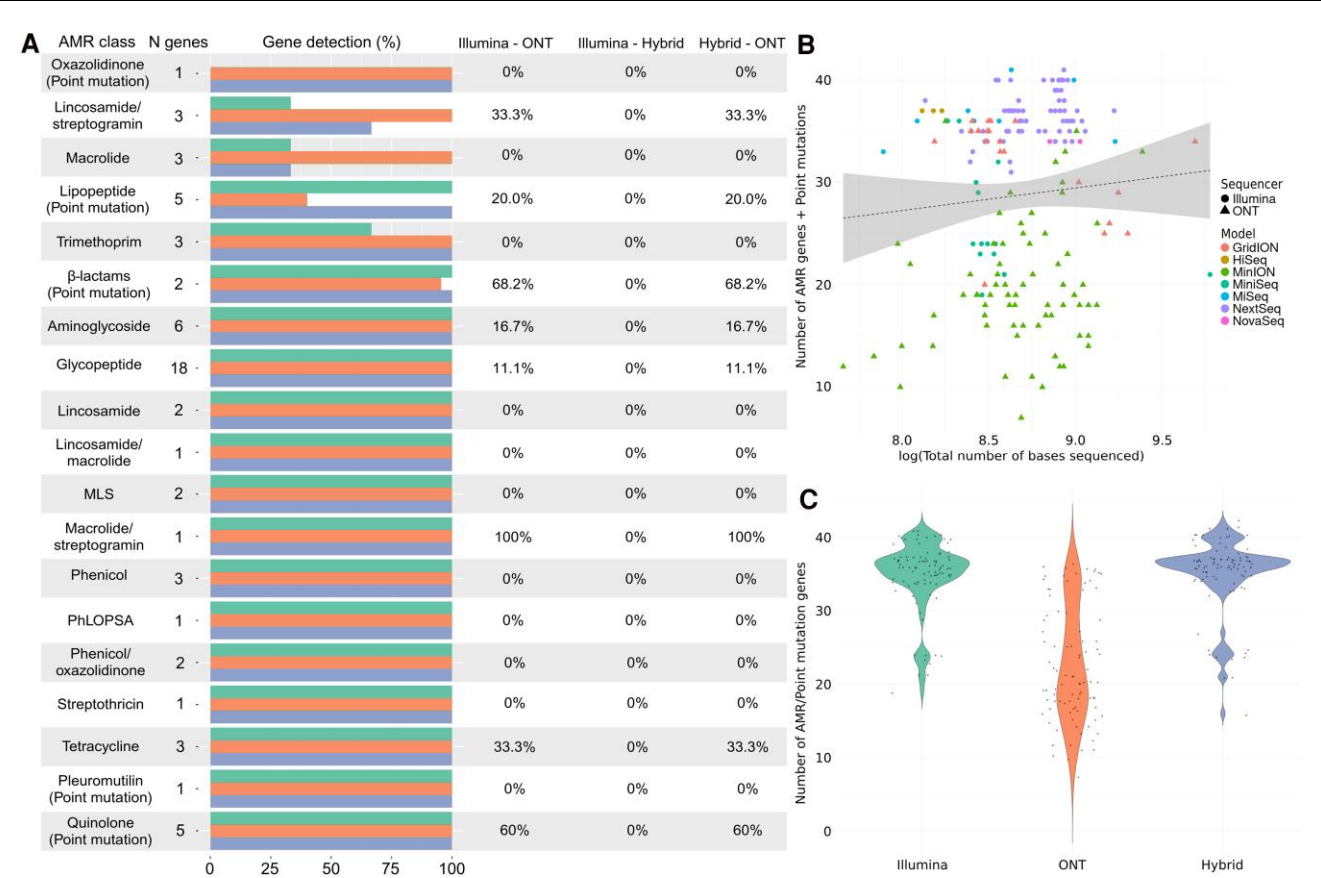


Figure 2. Comparative analysis of antimicrobial resistance (AMR) detection in *Enterococcus faecium* using different sequencing technologies. **A**, Differences in the identification of AMR genes and point mutations conferring resistance by Illumina (green), Oxford Nanopore Technologies (ONT; orange), and hybrid assembly (blue). The horizontal bars represent the total number and its corresponding percentage of genes detected within each antimicrobial class, followed by the percentage of genes that were significantly differently detected by each pair of sequencers (0% = no differences in detection). **B**, Relationship between the total number of sequenced bases by each sequencer model, from Illumina (HiSeq, MiniSeq, MiSeq, NextSeq, and NovaSeq) and ONT (GridION, MinION, and PromethION). **C**, Overall differences in the number of AMR and point mutation genes detected by Illumina and ONT sequencing platforms, where the y-axis represents the total number of genes and point mutations identified. Abbreviations: AMR, antimicrobial resistance; MLS, macrolide-lincosamide-streptogramin; ONT, Oxford Nanopore Technologies; PhLOPSA, Phenicol, Lincosamides, Oxazolidinones, Pleuromutilins, and Streptogramin A.

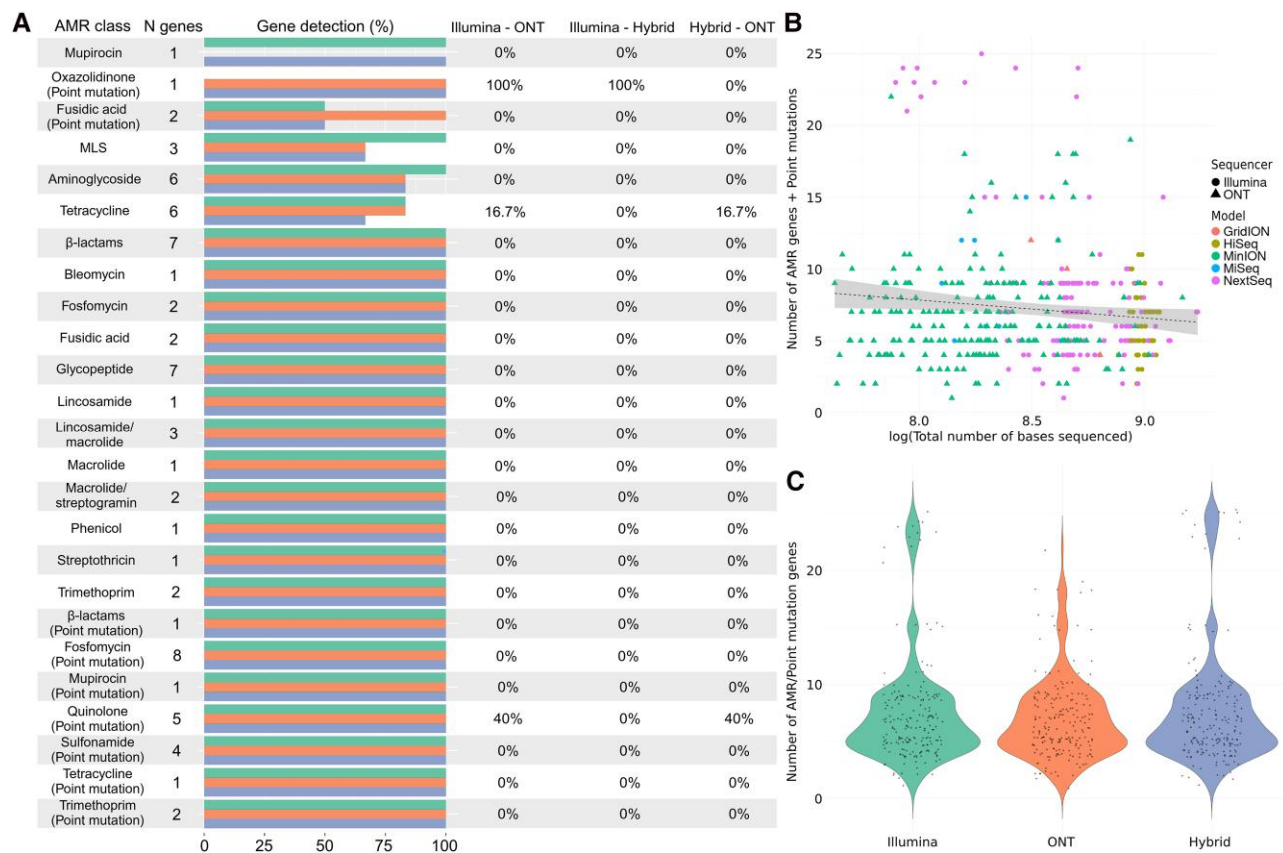


Figure 3. Comparative analysis of antimicrobial resistance (AMR) detection in *Staphylococcus aureus* using different sequencing technologies. **A**, Differences in the identification of AMR genes and point mutations conferring resistance by Illumina (green), Oxford Nanopore Technologies (ONT; orange), and hybrid assembly (blue). The horizontal bars represent the total number and its corresponding percentage of genes detected within each antimicrobial class, followed by the percentage of genes that were significantly differently detected by each pair of sequencers (0% = no differences in detection). **B**, Relationship between the total number of sequenced bases by each sequencer model, from Illumina (HiSeq, MiniSeq, MiSeq, NextSeq, and NovaSeq) and ONT (GridION, MinION, and PromethION). **C**, Overall differences in the number of AMR and point mutation genes detected by Illumina and ONT sequencing platforms, where the y-axis represents the total number of genes and point mutations identified. Abbreviations: AMR, antimicrobial resistance; MLS, macrolide-lincosamide-streptogramin; ONT, Oxford Nanopore Technologies.

sequenced per isolate, we did not find a specific pattern across the species. *Enterococcus faecium* and *A. baumannii* showed positive correlations between sequencing depth and the number of AMR genes and point mutations. In contrast, *S. aureus*, *K. pneumoniae*, *P. aeruginosa*, and *E. cloacae* showed a negative correlation between these variables. To further explore these results, we also analyzed the potential differences between AMR gene and point mutation detection within each sequencing technology by analyzing each platform as a covariate in analyses of covariance (ANCOVA). These results evidence the influence of sequencing depth in all species except for *S. aureus*, where these results were these differences were not statistically significant. Similarly, when we compared the overall detection of AMR genes and point mutations by each sequencing method, all species except for *S. aureus* showed significant differences, where hybrid correction identified the highest number, closely followed by Illumina and finally by ONT, which generally detected a significantly smaller number.

ESKAPE-Specific AMR Profiles. We found significant differences in AMR gene detection between the 3 methods, with hybrid correction detecting a higher number of genes, followed by Illumina and ONT, which identified a significantly fewer number (Figure 2A, Supplementary Table 3). Some examples of these differences were particularly clear in the case of aminoglycosides, glycopeptides, lincosamide/streptogramins, and tetracyclines (Figure 2, Supplementary Tables 2 and 3). In the case of point mutations, the isolates sequenced by Illumina exhibited the highest number, followed by hybrid assembly and finally by ONT (Supplementary Table 3). Overall, the highest number of point mutations detected were associated with β-lactam genes, followed by lipopeptide- and quinolone-associated genes (Figure 2, Supplementary Table 3). We found no significant association between sequencing depth of all *E. faecium* isolates used in this study and the overall combination of AMR genes plus point mutations ($P > .05$). However, we did not find significant differences

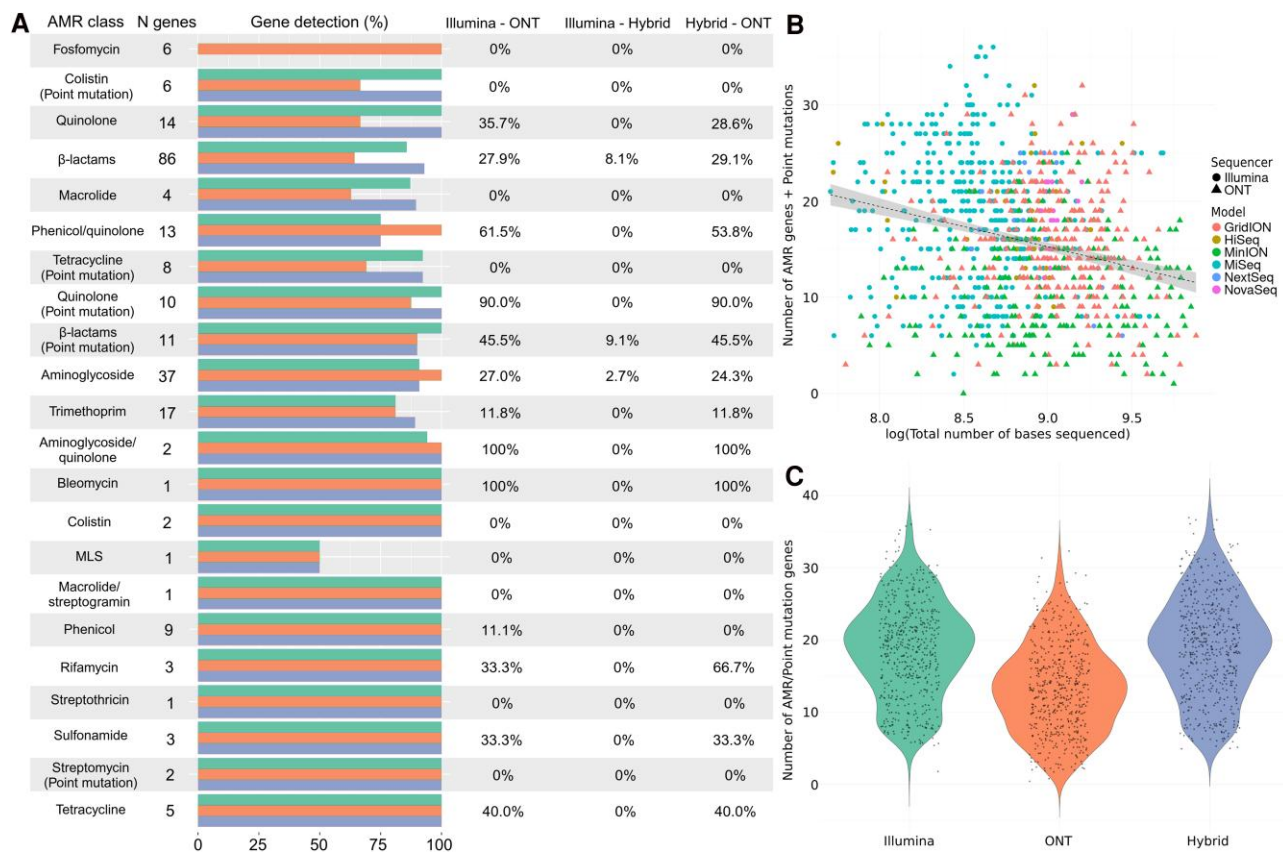


Figure 4. Comparative analysis of antimicrobial resistance (AMR) detection in *Klebsiella pneumoniae* using different sequencing technologies. **A**, Differences in the identification of AMR genes and point mutations conferring resistance by Illumina (green), Oxford Nanopore Technologies (ONT; orange), and hybrid assembly (blue). The horizontal bars represent the total number and its corresponding percentage of genes detected within each antimicrobial class, followed by the percentage of genes that were significantly differently detected by each pair of sequencers (0% = no differences in detection). **B**, Relationship between the total number of sequenced bases by each sequencer model, from Illumina (HiSeq, MiniSeq, MiSeq, NextSeq, and NovaSeq) and ONT (GridION, MinION, and PromethION). **C**, Overall differences in the number of AMR and point mutation genes detected by Illumina and ONT sequencing platforms, where the y-axis represents the total number of genes and point mutations identified. Abbreviations: AMR, antimicrobial resistance; MLS, macrolide-lincosamide-streptogramin; ONT, Oxford Nanopore Technologies.

between sequencer models (ie, MiSeq, NovaSeq, MinION, GridION, etc) when the sequencing technology was analyzed as a covariate (Illumina or ONT) following a positive association between number of genes and point mutations and sequencing depth (Figure 2B, Supplementary Table 4, ANCOVA results). Similarly, we did not find statistically significant differences in the overall number of AMR genes plus point mutations combined identified by each method (hybrid, Illumina, ONT) (Figure 2C, Supplementary Table 4, ANOVA results), with hybrid correction detecting the highest number.

The results for *S aureus* also showed that overall, hybrid assembly found the highest number of AMR genes, followed by Illumina and ONT (Supplementary Tables 2 and 3). These differences were significant for a number of antimicrobial classes such as β -lactams or glycopeptides (particularly the *Van* gene family). In the case of point mutations, hybrid assembly was able to detect a higher number than Illumina and ONT. These mutations were particularly present in genes associated

with fosfomycin, quinolone, and sulfonamide resistance (Figure 3A, Supplementary Tables 2 and 3). When comparing the relationship between sequencing depth of all *S aureus* isolates used in this study and the total number of AMR genes and point mutations, we found no significant differences (Figure 3B, Supplementary Table 4). Likewise, there were no differences on the overall number of AMR genes and point mutations identified, showing a consistent pattern across the 3 sequencing methods (Figure 3C, Supplementary Table 4).

We observed a wide variety of genes that showed significant differences in the detection of AMR genes depending on the sequencing method used for *K pneumoniae* sequences. It is important to note that *K pneumoniae* was the species with the highest number of analyzed isolates of the 6, which likely contributed to these results.

Klebsiella pneumoniae followed the same pattern as *E faecium*, where hybrid assembly detected the highest number of AMR genes followed by Illumina, with ONT identifying

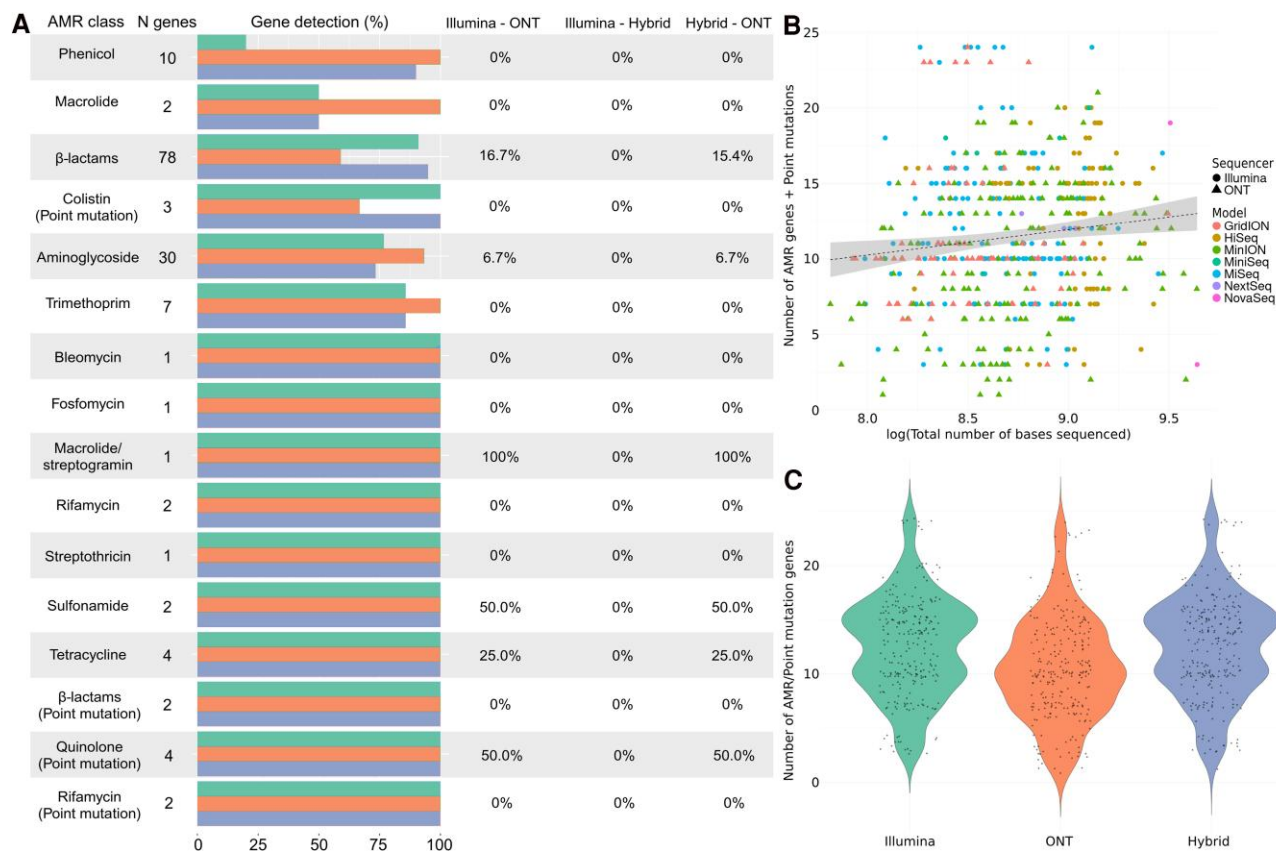


Figure 5. Comparative analysis of antimicrobial resistance (AMR) detection in *Acinetobacter baumannii* using different sequencing technologies. **A**, Differences in the identification of AMR genes and point mutations conferring resistance by Illumina (green), Oxford Nanopore Technologies (ONT; orange), and hybrid assembly (blue). The horizontal bars represent the total number and its corresponding percentage of genes detected within each antimicrobial class, followed by the percentage of genes that were significantly differently detected by each pair of sequencers (0% = no differences in detection). **B**, Relationship between the total number of sequenced bases by each sequencer model, from Illumina (HiSeq, MiniSeq, MiSeq, NextSeq, and NovaSeq) and ONT (GridION, MiniON, and PromethION). **C**, Overall differences in the number of AMR and point mutation genes detected by Illumina and ONT sequencing platforms, where the y-axis represents the total number of genes and point mutations identified. Abbreviations: AMR, antimicrobial resistance; ONT, Oxford Nanopore Technologies.

significantly fewer genes (Figure 4A, Supplementary Tables 2 and 3). Genes associated with carbapenem resistance (β -lactam class) presented the largest discrepancies between the sequencing methods, where hybrid assembly and Illumina showed similar levels of detection, both outperforming ONT. This same pattern was also observed for quinolone and aminoglycoside resistance, where ONT sequences identified significantly fewer genes than the other 2 methods. Hybrid assembly also showed a slightly higher detection of point mutations than Illumina, followed by ONT, which detected significantly fewer (Supplementary Table 3). These mutations were more often detected for β -lactam, quinolone, and tetracycline genes (Figure 4A, Supplementary Tables 2 and 3).

Klebsiella pneumoniae presented a strong significant negative relationship between the total number of bases sequenced of all isolates and the number of AMR genes and point mutations identified. These differences were observed both between sequencing technologies (Illumina vs ONT) and also when these technologies were assessed as a covariate (Figure 4B,

Supplementary Table 4). Similarly, we found significant differences between the number of AMR genes and point mutations identified by each of the methods, with hybrid assembly detecting a higher number than Illumina, followed by ONT (Figure 4C, Supplementary Table 4).

In the case of *A. baumannii*, hybrid assembly detected a slightly higher number of AMR genes, followed by Illumina and ONT. The difference between ONT with the other 2 approaches was particularly significant for antimicrobial classes such as β -lactams and aminoglycosides, which were the 2 most common in this species (Figure 5A, Supplementary Tables 2 and 3). In the case of point mutations, both hybrid assembly and Illumina identified an equal number of mutations, with ONT detecting significantly fewer (Supplementary Tables 2 and 3). A positive correlation between sequencing depth and the number of identified AMR genes and point mutations was observed in *A. baumannii* sequences both between platforms (Illumina vs ONT) and when each sequencing technology was analyzed as a covariate, although the correlation was

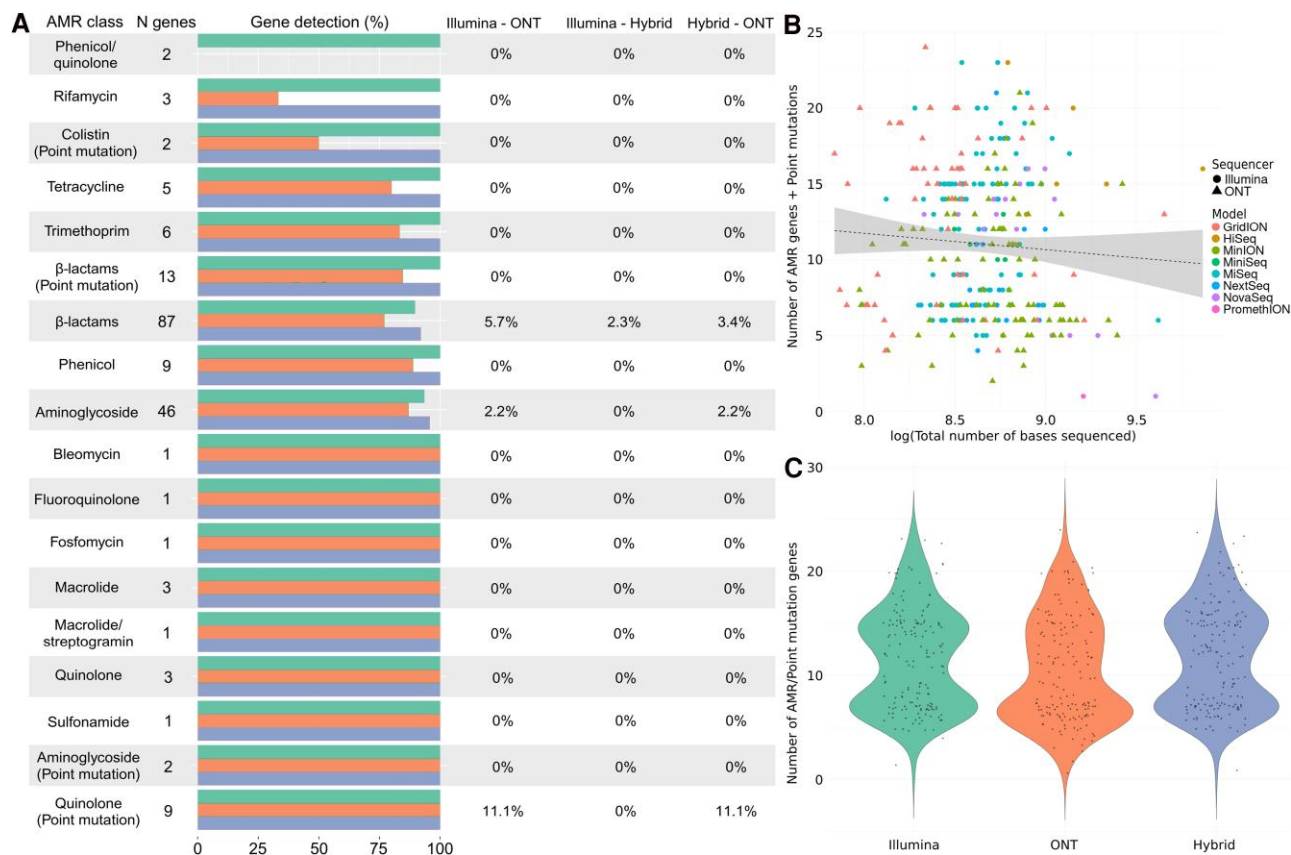


Figure 6. Comparative analysis of antimicrobial resistance (AMR) detection in *Pseudomonas aeruginosa* using different sequencing technologies. **A**, Differences in the identification of AMR genes and point mutations conferring resistance by Illumina (green), Oxford Nanopore Technologies (ONT; orange), and hybrid assembly (blue). The horizontal bars represent the total number and its corresponding percentage of genes detected within each antimicrobial class, followed by the percentage of genes that were significantly differently detected by each pair of sequencers (0% = no differences in detection). **B**, Relationship between the total number of sequenced bases by each sequencer model, from Illumina (HiSeq, MiniSeq, MiSeq, NextSeq, and NovaSeq) and ONT (GridION, MinION, and PromethION). **C**, Overall differences in the number of AMR and point mutation genes detected by Illumina and ONT sequencing platforms, where the y-axis represents the total number of genes and point mutations identified. Abbreviations: AMR, antimicrobial resistance; ONT, Oxford Nanopore Technologies.

weak (Figure 5B, Supplementary Table 4). Additionally, slight but significant differences were detected in the number of AMR genes identified by each method, with hybrid assembly detecting a higher number than Illumina or ONT (Figure 5C, Supplementary Table 4).

The detection of AMR genes for *P. aeruginosa* was higher for hybrid assembly, closely followed by Illumina and with ONT detecting fewer genes overall. Aminoglycoside and β -lactam resistance genes were the most prevalent, with the *aac* and *bla* gene families as the most detected (Figure 6A, Supplementary Tables 2 and 3). Similarly, point mutations associated with β -lactams were the most commonly found in this bacterium, followed by mutations associated with quinolones. As in other species, Illumina detected the highest number of point mutations, closely followed by hybrid assembly and ONT. There were no statistically significant associations between the overall sequencing depth of *P. aeruginosa* isolates and the number of AMR genes identified in *P. aeruginosa*. However, there was a

significant influence of sequencing deepness when sequencing platforms were measured as a covariate, showing a positive correlation (Figure 6B, Supplementary Table 4). Likewise, our results showed significant differences in the overall detection of AMR genes and point mutations between the 3 methods, with hybrid assembly identifying the highest number (Figure 6C, Supplementary Table 4).

Finally, *E. cloacae* presented significant differences in AMR gene detection between the methods, following the same pattern described previously, where hybrid correction detected a higher number of genes, followed by Illumina and finally ONT. Genes associated with aminoglycosides and β -lactams were the most prevalent, with the *aac* and *bla* gene families as the main groups of genes, respectively (Figure 7A, Supplementary Tables 2 and 3). *Enterobacter cloacae* only showed point mutations associated with quinolone resistance, particularly in the *gyrA* genes (Supplementary Tables 2 and 3), with Illumina detecting a higher number, followed by

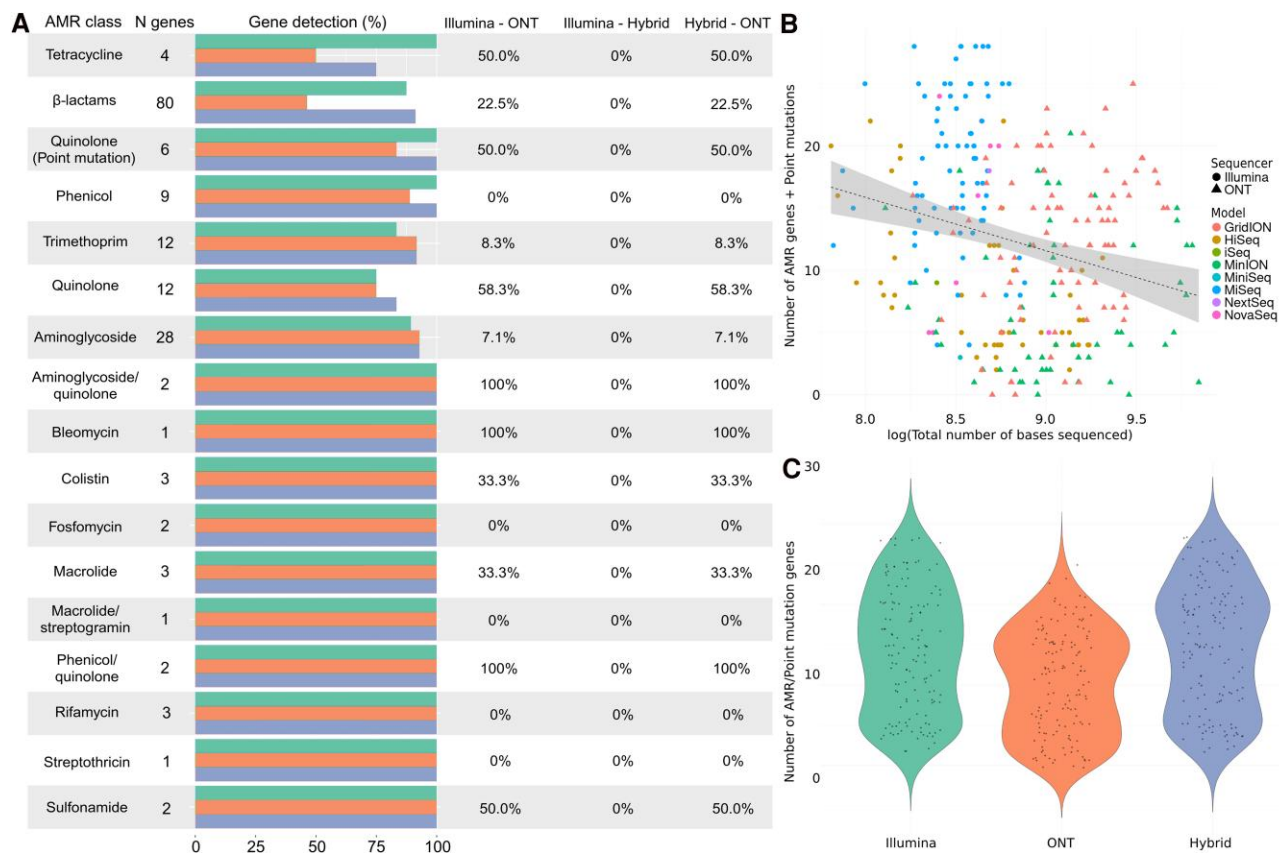


Figure 7. Comparative analysis of antimicrobial resistance (AMR) detection in *Enterobacter cloacae* using different sequencing technologies. **A**, Differences in the identification of AMR genes and point mutations conferring resistance by Illumina (green), Oxford Nanopore Technologies (ONT; orange), and hybrid assembly (blue). The horizontal bars represent the total number and its corresponding percentage of genes detected within each antimicrobial class, followed by the percentage of genes that were significantly differently detected by each pair of sequencers (0% = no differences in detection). **B**, Relationship between the total number of sequenced bases by each sequencer model, from Illumina (HiSeq, MiniSeq, MiSeq, NextSeq, and NovaSeq) and ONT (GridION, MinION, and PromethION). **C**, Overall differences in the number of AMR and point mutation genes detected by Illumina and ONT sequencing platforms, where the y-axis represents the total number of genes and point mutations identified. Abbreviations: AMR, antimicrobial resistance; ONT, Oxford Nanopore Technologies.

hybrid assembly and ONT. *Enterobacter cloacae* presented a significant negative relationship between the total number of bases sequenced and the number of identified AMR genes and point mutations both when all sequences were analyzed together and when each platform was analyzed as a covariate (Figure 7B, Supplementary Table 4). This bacterium also showed significant differences between the number of AMR genes and point mutations identified by the 3 methods, with hybrid assembly detecting the highest number, followed by Illumina and ONT (Figure 7C, Supplementary Table 4).

Pangenome Characterization

Generally, a consistent trend emerged in identifying genes across all 6 species. Among the 3 sequencing methods employed, Illumina-sequenced isolates consistently exhibited the lowest number of detected genes, while ONT-sequenced counterparts demonstrated the highest count. The hybrid assembly method typically yielded gene counts falling between those

obtained from Illumina and ONT. This pattern can be observed in panel A of Supplementary Figures 2–7, which visualizes the comparison of the number of genes in the core genome detected by each technology (Illumina, ONT, and hybrid correction).

However, species-specific proportions of cloud, core, soft-core, and shell genes did not exhibit discernible patterns associated with the sequencing type but were instead associated with the bacterial species (Table 2). The proportion of genes per pangenome structure (core, soft-core, shell, and cloud) detected by each method is available in panel B of Supplementary Figures 2–7. A recurrent observation across the species was that Illumina sequencing consistently revealed fewer unique accessory genes than the other methods. This can be observed by looking at the pangenome rarefaction curves (panel C of Supplementary Figures 2–7). Pangenome rarefaction curves are graphical visualizations of the growth of the pangenome size as a function of the number of genomes (or isolates) that are included in the analysis. Each point indicates the number

Table 2. Summary of the Pangenome Gene Classification in ESKAPE Pathogens

Species	Sequencer	Core Genes	Soft-Core Genes	Shell Genes	Cloud Genes	Total Genes, No.
<i>Enterococcus faecium</i>	Illumina	1706 (27.5)	288 (4.6)	1425 (22.9)	2795 (45.0)	6214
	ONT	1529 (8.5)	387 (2.1)	3339 (18.5)	12 793 (70.9)	18 048
	Hybrid	1696 (11.82)	300 (2.09)	2018 (14.06)	10 337 (72.03)	14 351
<i>Staphylococcus aureus</i>	Illumina	2117 (27.1)	55 (0.7)	842 (10.8)	4802 (61.4)	7816
	ONT	2039 (27.5)	90 (1.2)	1013 (13.7)	4278 (57.7)	7420
	Hybrid	2115 (36.47)	45 (0.78)	884 (15.24)	2755 (47.51)	5799
<i>Klebsiella pneumoniae</i>	Illumina	978 (2.6)	3097 (8.1)	2139 (5.6)	31 974 (83.7)	38 188
	ONT	1888 (2.3)	1548 (1.9)	7702 (9.3)	72 109 (86.6)	83 247
	Hybrid	3201 (7.24)	944 (2.14)	2287 (5.17)	37 764 (85.45)	44 196
<i>Acinetobacter baumannii</i>	Illumina	1904 (11.7)	763 (4.7)	1651 (10.2)	11 925 (73.4)	16 243
	ONT	2038 (5.74)	580 (1.63)	2999 (8.45)	29 859 (84.17)	35 476
	Hybrid	2494 (13.9)	290 (1.6)	1675 (9.3)	13 499 (75.2)	17 958
<i>Pseudomonas aeruginosa</i>	Illumina	53 (0.1)	4981 (13.2)	2177 (5.8)	30 447 (80.9)	37 658
	ONT	150 (0.3)	3953 (8.2)	4331 (9.0)	39 748 (82.5)	48 182
	Hybrid	256 (0.77)	4901 (14.75)	2209 (6.65)	25 859 (77.83)	33 225
<i>Enterobacter cloacae</i>	Illumina	1004 (2.8)	1947 (5.5)	2900 (8.2)	29 676 (83.5)	35 527
	ONT	1719 (2.4)	737 (1.0)	5020 (7.1)	62 938 (89.4)	70 414
	Hybrid	2556 (6.4)	621 (1.6)	2767 (6.9)	33 686 (85.0)	39 630

Data are presented as No. (%) unless otherwise indicated.

Abbreviation: ONT, Oxford Nanopore Technologies.

of genes present in a specific genome. Therefore, genes that are not identical between 2 samples are identified as 2 individual genes, even if they are the same gene variants. As the number of analyzed genomes increases, so does the value found in the x-axis.

The shape of these curves represents the open or closed nature of the pangenome, where a continuously rising curve corresponds to an open pangenome, meaning that as new samples are included significant numbers of genes are added to the pangenome. On the other hand, if the rarefaction curve reaches a plateau, it indicates a closed pangenome since adding new isolates does not increase the pangenome size. In our case, Illumina sequences tend to lead to a narrower shape of the rarefaction curves, showing that this technology tends to find more genomic similarities and, therefore, fewer unique gene variants between the isolates (blue curve in panel C in [Supplementary Figures 2–7](#)). This suggests a less diverse group of isolates within each species, except for *P. aeruginosa*, where this particular pattern was not observed.

For *E. faecium*, ONT identified the highest number of core genes, followed by hybrid correction and Illumina ([Supplementary Figure 2A](#)), which exhibited a reduced proportion of accessory genome compared to ONT or hybrid assembly. The latter displayed a pattern more similar to ONT results ([Supplementary Figure 2B](#)). These distinctions are observable in the pangenome rarefaction curves, where ONT and hybrid assembly reveal a higher share of distinctive accessory genes ([Supplementary Figure 2C](#)).

Conversely, *S. aureus* isolates sequenced using Illumina showed a closer number of core genes to those detected by

ONT, with most genes detected by hybrid correction being shared between the methods ([Supplementary Figure 3A](#)). Following this pattern, Illumina presented a more similar percentage of cloud genes than ONT and hybrid correction, with hybrid correction showing the lowest proportion of cloud genes of the 3 methods ([Supplementary Figure 3B](#)). Finally, this bacterium's rarefaction curves were more similar, with ONT still presenting a slightly wider shape ([Supplementary Figure 3C](#)).

For *K. pneumoniae*, ONT detected a distinctively higher number of core genes compared to the other 2 methods, which were more similar when compared ([Supplementary Figure 4A](#)). There were no significant variations in the proportion of cloud genes identified by each method. Nevertheless, Illumina exhibited the greatest ratio of soft-core genes, ONT displayed the highest percentage of shell genes, and with hybrid correction, there was a comparable distribution between core and shell genes, while the number of soft-core genes was the lowest among all methods ([Supplementary Figure 4B](#)). This was observed in the rarefaction curves, where hybrid correction displayed the widest shape of the 3 methods ([Supplementary Figure 4C](#)).

The findings for *A. baumannii* resemble those of *E. faecium* and *K. pneumoniae*. ONT revealed the highest number of core genes ([Supplementary Figure 5A](#)), and the highest percentage of cloud genes, while the percentage of shell genes remained consistent across all 3 options. Ultimately, Illumina displayed the highest proportion of soft-core genes ([Supplementary Figure 5B](#)). The shape of the rarefaction curves was more similar between Illumina and hybrid correction, with ONT presenting a slightly wider distribution ([Supplementary Figure 5C](#)).

Pseudomonas aeruginosa stood out as the sole species with an almost negligible core genome (0.1%–0.7%), showing a clear scattered pangenome curve when compared to the other species. For this bacterium, the number of core genes was similar between the 3 methods, with ONT still detecting a higher number of genes (Supplementary Figure 6A). Hybrid correction yielded fewer cloud genes, whereas Illumina and ONT exhibited comparable percentages. Regarding shell genes, ONT sequences showed the highest proportion, followed by the hybrid correction and Illumina. Notably, the proportion of soft-core genes was similar between Illumina and hybrid correction, both surpassing the levels detected in ONT (Supplementary Figure 6B). The rarefaction curves for this species were generally spread in shape, with Illumina being the wider of the 3 (Supplementary Figure 6C).

Enterobacter cloacae followed similar patterns to *K pneumoniae*, with ONT detecting the highest number of core genes compared to either Illumina or hybrid correction (Supplementary Figure 7A). A comparable proportion of shell genes was identified by all 3 methods. Illumina sequences exhibited the highest percentage of soft-core genes, whereas hybrid correction showed the highest percentage of core genes. Ultimately, ONT and the hybrid approach identified the highest proportion of cloud genes (Supplementary Figure 7B). Finally, the shape of the rarefaction curves for this species was similar between the 3 methods, with hybrid correction presenting a slightly wider distribution (Supplementary Figure 7C).

The results of the ANOVA tests comparing the number of annotated genes prior to pangenome reconstruction showed statistically significant differences between the 3 approaches (Illumina, ONT, and hybrid assembly) across all 6 bacterial species. Illumina sequences consistently presented fewer annotated genes compared to both ONT and hybrid assembly, but these 2 did not show statistical differences between each other in any of the bacteria (Supplementary Table 5, Supplementary Figure 8).

DISCUSSION

This comprehensive study was designed to compare the outcomes of similar bioinformatic analyses conducted on isolates from ESKAPE pathogens (*E faecium*, *S aureus*, *K pneumoniae*, *A baumannii*, *P aeruginosa*, and *E cloacae*) that were sequenced with both short-read (Illumina) and long-read (ONT) platforms, as well as the emerging hybrid assembly approach. While Illumina and ONT sequencing effectively identified key AMR genes, our findings revealed significant variability across the different methods.

The differences in performance between ONT and Illumina can be partially attributed to the fact that most bioinformatic tools have historically been developed to process short-read sequences due to the longer presence and accessibility of Illumina technology [66, 67]. Consequently, these tools have had more

time to refine their pipelines and algorithms for short-read data analysis, which may not be optimized for the distinct characteristics of ONT long-read sequences. This disadvantage was observed in our genome completeness (BUSCO) results, which showed better performance on Illumina sequences compared to ONT. Despite discrepancies in our BUSCO results, we included these findings to provide complementary information supporting the observed differences between Illumina and ONT sequencers. BUSCO scoring of predicted genes is based on hidden Markov model (HMM) profiles that must meet a score threshold [68]. BUSCO's gene predictor can encounter spurious stop codons caused by the high error rate of the long-read technology. As a result, a truncated protein is passed to HMM to be scored, which will lower BUSCO results. With Illumina, the error rate is lower, so these early stop codons are not as common (M. Seppey, personal communication, BUSCO developer). We observed a substantial improvement in assessing genome completeness when applying the hybrid assembly method, which also led to an increase in the number of AMR genes identified, nearing what we found in the isolates sequenced with Illumina. These results demonstrate how discrepancies in tool development timelines underscore the need for further advancements in bioinformatics tailored to long-read data analysis to fully harness the potential of ONT technology in genomics and molecular research, highlighting the need for improved tools for long reads.

Our results align with the previously described AMR repertoire for ESKAPE pathogens, with a high prevalence of genes conferring resistance to β -lactams throughout most species [69]. β -lactam genes are known to present a wide variety of closely related variants [70, 71], where short-read sequencing tends to have higher accuracy at individual-base level than long-read sequencing, due to its better performance for single-nucleotide polymorphism analysis [13, 72]. The existence of these variants is likely the main reason behind the difference in the number of observed β -lactam genes between Illumina and ONT sequences, especially in species known to present a higher number of β -lactam gene variants such as *A baumannii*, *E cloacae*, and *P aeruginosa*, the first 2 known for their complex resistomes [73–75]. Genes conferring resistance against aminoglycosides, macrolides, and tetracyclines were consistently identified by both sequencers across all species, and therefore also by hybrid correction. This similarity indicates that current tools and databases for detecting AMR genes perform well across platforms. Resistance against other antimicrobial classes was more variable, independent of the sequencing method.

Similarly, our results also align with previous studies describing that Illumina sequencers tend to outperform ONT sequencers at point mutation detection in other organisms [23, 76]. These differences primarily arise from the inherent error profiles associated with read length of each sequencing platform, positioning emerging approaches such as hybrid assembly as

a great option to improve ONT accuracy while more robust downstream analysis pipelines are established [31].

We did not observe a conserved pattern showing the directionality of the relationship between overall sequencing depth and AMR gene and point mutation detection. However, the influence of sequencing deepness was evident in all species except for *S aureus*, which presented comparable results across the 3 methods. These similarities could be due to the clinical importance of *S aureus* and the detailed characterization of its genome, but also due to a potential higher similarity between the isolates used for the analyses.

This difference between the species was also observed in the characterization of the pangenome size and its composition, where *S aureus* presented minimal variation between the 3 methods compared to the rest of ESKAPE pathogens. These differences most likely arise from variations in gene annotation, where isolates sequenced by Illumina consistently resulted in significantly fewer annotated genes compared to either ONT or hybrid assembly. These results indicate that the choice of sequencing technology can have a significant influence in gene annotation in bacterial genome assemblies, which may have potential implications for downstream analyses as pangenome reconstruction.

Thus, our results highlight that the use of publicly available data from short- and long-read sequencing technologies can present enough variations in the obtained outcomes that the decision of using either long or short sequences, or their combination through hybrid assembly should be driven by the specific goals, bacterial species, genes of interest, and scope of the study. This is particularly important for studies that use publicly available genomes that were sequenced across a relatively wide timespan, since the differences between these technologies were more notorious on the early stages of long-read sequencing. For instance, our results support previous literature that suggests Illumina as a great option for common AMR analysis and pangenome reconstruction, due to its per-base accuracy, which is crucial for confidently identifying resistance [77], while ONT can be a better option for the study of newly discovered bacterial species with potentially rearranged genomes since it is able to generate reads spanning tens of kilobases, offering a clearer view of large insertions or deletions, which in the case of pangenome reconstruction, will provide a comprehensive coverage of complex regions [78, 79]. Our observations on the application of hybrid assembly also prove its potential as a promising strategy to optimize the benefits of both technologies for a more comprehensive understanding of bacterial pathogens' genetic makeup [38, 80].

Finally, it is noteworthy that our analyses were based on isolates for which both short- and long-read sequences were already available, which may be an indicator of specific objectives on their original studies that could potentially bring bias of the type of isolates collected. For that reason, these isolates may not be accurate representations of the actual variation

and diversity of the resistome in ESKAPE pathogens. One likely example of this is the unusual low resistance found for *S aureus* isolates compared to previously described information. Therefore, the results of the current study must be interpreted aiming to identify and understand the differences that can be encountered in the obtained results depending on the use of short reads (Illumina), long reads (ONT), or hybrid assembly for the study of AMR, AMR-associated point mutations, or pangenome composition and structure, and serve as a guide for a data-driven decision for future research.

CONCLUSIONS

Our study compared short-read (Illumina) and long-read (ONT) sequencing technologies and the application of hybrid assembly, for analyzing AMR-related gene detection in ESKAPE pathogens at the pangenome level. Both technologies identified key genetic elements; hybrid assembly often outperformed both individual methods for AMR gene detection, closely followed by Illumina and finally by ONT, which detected significantly fewer genes. However, in the case of point mutation detection, Illumina was able to detect a higher number than the other 2 methods. Genome completeness was higher for short-read sequences, likely due to the existence of optimized bioinformatic tools. While no consistent pattern of correlation between AMR detection and sequencing depth emerged across species, this variable significantly influenced results within individual platforms. Pangenome analysis showed that ONT can identify a higher number of core and accessory genes, while Illumina found fewer unique variants between isolates likely due to the consistently lower number of annotated genes detected by it. For these reasons, this study highlights the need to take both pathogen and specific goals of the study into consideration when selecting a sequencing approach particularly when using publicly available genomic data ranging a wide timespan. This study provides valuable insights into the strengths, limitations, and synergies of different sequencing technologies, guiding researchers in making informed choices for comprehensive genetic analyses of bacterial pathogens.

Supplementary Data

[Supplementary materials](#) are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Author contributions. A. F.-D.-D. and M. J.: Conception and design of work; acquisition, analysis, and interpretation of data; and writing and revision of the manuscript. C. L.: Conception and design of work; interpretation of data; writing and revision of the manuscript; acquisition of funding; and supervision of project. All authors have approved the submitted version of the manuscript and have agreed to be accountable for their contribution to the article.

Disclaimer. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funder.

Financial support. This material is based upon work supported by the US National Institutes of Health (NIH) (grant number R35GM134934).

Potential conflicts of interest. The authors: No reported conflicts of interest.

References

- Murray CJ, Ikuta KS, Sharara F, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* **2022**; 399:629–55.
- Mulani MS, Kamble EE, Kumkar SN, Tawre MS, Pardesi KR. Emerging strategies to combat ESKAPE pathogens in the era of antimicrobial resistance: a review. *Front Microbiol* **2019**; 10:539.
- Rice LB. Federal funding for the study of antimicrobial resistance in nosocomial pathogens: no ESKAPE. Chicago, IL: University of Chicago Press, **2008**:1079–81.
- Ma YX, Wang CY, Li YY, et al. Considerations and caveats in combating ESKAPE pathogens against nosocomial infections. *Adv Sci* **2020**; 7:1901872.
- Vale de Macedo GHR, Costa GDE, Oliveira ER, et al. Interplay between ESKAPE pathogens and immunity in skin infections: an overview of the major determinants of virulence and antibiotic resistance. *Pathogens* **2021**; 10:148.
- Goel V, Kumar D, Kumar R, Mathur P, Singh S. Community acquired enterococcal urinary tract infections and antibiotic resistance profile in north India. *J Lab Physicians* **2016**; 8:050–4.
- Ruan Z, Yu Y, Feng Y. The global dissemination of bacterial infections necessitates the study of reverse genomic epidemiology. *Brief Bioinform* **2020**; 21:741–50.
- Sherry NL, Lane CR, Kwong JC, et al. Genomics for molecular epidemiology and detecting transmission of carbapenemase-producing Enterobacteriales in Victoria, Australia, 2012 to 2016. *J Clin Microbiol* **2019**; 57:e00573–19.
- Crispell J, Benton CH, Balaz D, et al. Combining genomics and epidemiology to analyse bi-directional transmission of *Mycobacterium bovis* in a multi-host system. *Elife* **2019**; 8:e45833.
- Yelin I, Flett KB, Merakou C, et al. Genomic and epidemiological evidence of bacterial transmission from probiotic capsule to blood in ICU patients. *Nat Med* **2019**; 25:1728–32.
- Vincent AT, Schietekatte O, Goarant C, et al. Revisiting the taxonomy and evolution of pathogenicity of the genus *Leptospira* through the prism of genomics. *PLoS Negl Trop Dis* **2019**; 13:e0007270.
- Zhu Q, Mai U, Pfeiffer W, et al. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nat Commun* **2019**; 10:5477.
- Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: an overview. *Hum Immunol* **2021**; 82:801–11.
- Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol* **2011**; 12:1–14.
- Whiteford N, Haslam N, Weber G, et al. An analysis of the feasibility of short read sequencing. *Nucleic Acids Res* **2005**; 33:e171.
- Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* **2020**; 21:1–16.
- Mantere T, Kersten S, Hoischen A. Long-read sequencing emerging in medical genetics. *Front Genet* **2019**; 10:426.
- Schuster CF, Weber RE, Weig M, Werner G, Pfeifer Y. Ultra-deep long-read sequencing detects IS-mediated gene duplications as a potential trigger to generate arrays of resistance genes and a mechanism to induce novel gene variants such as bla CTX-M-243. *J Antimicrob Chemother* **2022**; 77:381–90.
- Das S, Bombaywala S, Srivastava S, Kapley A, Dhodapkar R, Dafale NA. Genome plasticity as a paradigm of antibiotic resistance spread in ESKAPE pathogens. *Environ Sci Pollut Res Int* **2022**; 29:40507–19.
- Da Silva GJ, Domingues S. Insights on the horizontal gene transfer of carbapenemase determinants in the opportunistic pathogen *Acinetobacter baumannii*. *Microorganisms* **2016**; 4:29.
- Delahaye C, Nicolas J. Sequencing DNA with nanopores: troubles and biases. *PLoS One* **2021**; 16:e0257521.
- Riaz N, Leung P, Barton K, et al. Adaptation of Oxford Nanopore technology for hepatitis C whole genome sequencing and identification of within-host viral variants. *BMC Genomics* **2021**; 22:1–12.
- Tshiabula D, Gandhari J, Pillay S, et al. Comparison of SARS-CoV-2 sequencing using the ONT GridION and the Illumina MiSeq. *BMC Genomics* **2022**; 23:1–17.
- Dohm JC, Peters P, Stralis-Pavese N, Himmelbauer H. Benchmarking of long-read correction methods. *NAR Genom Bioinform* **2020**; 2:lqaa037.
- Tvedte ES, Gasser M, Sparklin BC, et al. Comparison of long-read sequencing technologies in interrogating bacteria and fly genomes. *G3 (Bethesda)* **2021**; 11:jkab083.
- Wick RR, Judd LM, Wyres KL, Holt KE. Recovery of small plasmid sequences via Oxford Nanopore sequencing. *Microb Genom* **2021**; 7:000631.
- Sauvage T, Cormier A, Delphine P. A comparison of Oxford Nanopore Library strategies for bacterial genomics. *BMC Genomics* **2023**; 24:627.
- Ruekit S, Srijan A, Serichantalergs O, et al. Molecular characterization of multidrug-resistant ESKAPEE pathogens from clinical samples in Chonburi, Thailand (2017–2018). *BMC Infect Dis* **2022**; 22:1–12.
- Ternus KL, Keplinger NC, Kappell AD, et al. Detection of ESKAPE pathogens and *Clostridioides difficile* in simulated skin transmission events with metagenomic and metatranscriptomic sequencing. *bioRxiv* [Preprint]. Posted online 4 March 2021. doi:10.1101/2021.03.04.433847
- Yadav S, Kapley A. Exploration of activated sludge resistome using metagenomics. *Sci Total Environ* **2019**; 692:1155–64.
- Maboni G, Baptista R, Wireman J, Framst I, Summers AO, Sanchez S. Three distinct annotation platforms differ in detection of antimicrobial resistance genes in long-read, short-read, and hybrid sequences derived from total genomic DNA or from purified plasmid DNA. *Antibiotics* **2022**; 11:1400.
- Kimani R, Musundi S, Wakaba P, et al. Application of nanopore sequencing to identify antimicrobial resistance genes, mobile genetic elements and virulence factors in clinical isolates. *bioRxiv* [Preprint]. Posted online 26 October 2023. doi:10.1101/2023.10.25.563907
- Conzemius R, Bergman Y, Májek P, et al. Automated antimicrobial susceptibility testing and antimicrobial resistance genotyping using Illumina and Oxford Nanopore Technologies sequencing data among Enterobacteriaceae. *Front Microbiol* **2022**; 13:973605.
- Judge K, Harris SR, Reuter S, Parkhill J, Peacock SJ. Early insights into the potential of the Oxford Nanopore MinION for the detection of antimicrobial resistance genes. *J Antimicrob Chemother* **2015**; 70:2775–8.
- Wu X, Luo H, Ge C, et al. Evaluation of multiplex nanopore sequencing for *Salmonella* serotype prediction and antimicrobial resistance gene and virulence gene detection. *Front Microbiol* **2023**; 13:1073057.
- Weinmaier T, Conzemius R, Bergman Y, et al. Validation and application of long-read whole-genome sequencing for antimicrobial resistance gene detection and antimicrobial susceptibility testing. *Antimicrob Agents Chemother* **2023**; 67:e01072–22.
- Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **2020**; 36:2253–5.
- De Maio N, Shaw LP, Hubbard A, et al. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microb Genom* **2019**; 5:e000294.
- Hyun JC, Monk JM, Palsom BO. Comparative pangenomics: analysis of 12 microbial pathogen pangenomes reveals conserved global structures of genetic and functional diversity. *BMC Genomics* **2022**; 23:1–18.
- Golchha NC, Nighojkar A, Nighojkar S. Redefining genomic view of *Clostridioides difficile* through pangenome analysis and identification of drug targets from its core genome. *Drug Target Insights* **2022**; 16:17.
- Zhou Y, Jiang D, Yao X, et al. Pan-genome wide association study of *Glaesserella parasuis* highlights genes associated with virulence and biofilm formation. *Front Microbiol* **2023**; 14:1160433.
- Zakham F, Sironen T, Vapalahti O, Kant R. Pan and core genome analysis of 183 *Mycobacterium tuberculosis* strains revealed a high inter-species diversity among the human adapted strains. *Antibiotics* **2021**; 10:500.
- Yang T, Zhong J, Zhang J, et al. Pan-genomic study of *Mycobacterium tuberculosis* reflecting the primary/secondary genes, generality/individuality, and the inter-conversion through copy number variations. *Front Microbiol* **2018**; 9:1886.
- Norsigian CJ, Danhof HA, Brand CK, et al. Systems biology approach to functionally assess the *Clostridioides difficile* pangenome reveals genetic diversity with discriminatory power. *Proc Natl Acad Sci U S A* **2022**; 119:e2119396119.
- Fatoba AJ, Fatoba DO, Babalola SO. Pangenome and subtractive genomic analysis of *Clostridioides difficile* reveals putative drug targets. *J Proteins Proteom* **2022**; 13:247–56.
- Katz K, Shutov O, Lapoint R, et al. The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Res* **2022**; 50(D1):D387–90.
- Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* **2021**; 38:4647–54.
- Petit RA III, Read TD. Bactopia: a flexible pipeline for complete analysis of bacterial genomes. *mSystems* **2020**; 5:e00190–20.
- Andrews S. FastQC: a quality control tool for high throughput sequence data. Cambridge, UK: Babraham Bioinformatics, Babraham Institute, **2010**.

50. Bushnell B. BB map short read aligner, and other bioinformatic tools. **2020**. Available at: [sourceforge net/projects/bbmap](https://sourceforge.net/projects/bbmap). Accessed December, 2024.
51. Petit RA III. Fastq-scan: generate summary statistics of input FASTQ sequences. **2022**. Available at: <https://github.com/rpetit3/fastq-scan>.
52. De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* **2023**; 39:btad311.
53. Steinig E, Coin L. Nanoq: ultra-fast quality control for nanopore reads. *J Open Source Softw* **2022**; 7:2991.
54. Bonenfant Q, Noé L, Touzet H. Porechop_ABI: discovering unknown adapters in Oxford Nanopore Technology sequencing reads for downstream trimming. *Bioinform Adv* **2022**; 3:vbac085.
55. Hall MB. Rasusa: randomly subsample sequencing reads to a specified coverage. *J Open Source Softw* **2022**; 7:3941.
56. Seemann T. Shovill: faster SPAdes assembly of Illumina reads. **2017**. <https://github.com/tseemann/shovill?tab=readme-ov-file>. Accessed December, 2024.
57. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **2012**; 19:455–77.
58. Ondov BD, Starrett GJ, Sappington A, et al. Mash Screen: high-throughput sequence containment estimation for genome discovery. *Genome Biol* **2019**; 20: 1–13.
59. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **2013**; 29:1072–5.
60. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **2014**; 30:2068–9.
61. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* **2017**; 13: e1005595.
62. Feldgarden M, Brover V, Haft DH, et al. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob Agents Chemother* **2019**; 63: e00483-19.
63. Zankari E, Allesøe R, Joensen KG, Cavaco LM, Lund O, Aarestrup FM. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *J Antimicrob Chemother* **2017**; 72:2764–8.
64. R Core Team. R: a language and environment for statistical computing. V 4.4.3. Vienna, Austria: R Foundation for Statistical Computing, **2020**.
65. Tonkin-Hill G, MacAlasdair N, Ruis C, et al. Producing polished prokaryotic pan-genomes with the Panaroo pipeline. *Genome Biol* **2020**; 21:1–21.
66. Moti TB, Guyassa C, Dima C, Home M. Illumina sequencing technology review. *Image* **2022**; 10:25–31.
67. Eisenstein M. Oxford Nanopore announcement sets sequencing sector abuzz. *Nat Biotechnol* **2012**; 30:295–6.
68. Lottaz C, Iseli C, Jongeneel CV, Bucher P. Modeling sequencing errors by combining hidden Markov models. *Bioinformatics* **2003**; 19(Suppl 2):ii103–i12.
69. De Oliveira DM, Forde BM, Kidd TJ, et al. Antimicrobial resistance in ESKAPE pathogens. *Clin Microbiol Rev* **2020**; 33:e00181-19.
70. Chewapreecha C, Marttinen P, Croucher NJ, et al. Comprehensive identification of single nucleotide polymorphisms associated with beta-lactam resistance within pneumococcal mosaic genes. *PLoS Genet* **2014**; 10:e1004547.
71. Rocha GD, Nogueira JF, Dos Santos MVG, et al. Impact of polymorphisms in *blaZ*, *blaR1* and *blaI* genes and their relationship with β -lactam resistance in *S. aureus* strains isolated from bovine mastitis. *Microb Pathog* **2022**; 165:105453.
72. Linde J, Brangsch H, Hölzer M, et al. Comparison of Illumina and Oxford Nanopore Technology for genome analysis of *Francisella tularensis*, *Bacillus anthracis*, and *Brucella suis*. *BMC Genomics* **2023**; 24:258.
73. Rodrigues DLN, Morais-Rodrigues F, Hurtado R, et al. Pan-resistome insights into the multidrug resistance of *Acinetobacter baumannii*. *Antibiotics* **2021**; 10: 596.
74. Annajhala MK, Gomez-Simmonds A, Uhlemann A-C. Multidrug-resistant *Enterobacter cloacae* complex emerging as a global, diversifying threat. *Front Microbiol* **2019**; 10:44.
75. Sethuvel DPM, Bakthavatchalam YD, Karthik M, et al. β -lactam resistance in ESKAPE pathogens mediated through modifications in penicillin-binding proteins: an overview. *Infect Dis Ther* **2023**; 12:829–41.
76. Garcia-Pedemonte D, Carcereny A, Gregori J, et al. Comparison of Nanopore and synthesis-based next-generation sequencing platforms for SARS-CoV-2 variant monitoring in wastewater. *Int J Mol Sci* **2023**; 24:17184.
77. Rose R, Nolan DJ, Ashcraft D, et al. Comparing antimicrobial resistant genes and phenotypes across multiple sequencing platforms and assays for Enterobacterales clinical isolates. *BMC Microbiol* **2023**; 23:225.
78. Rausch T, Snajder R, Leger A, et al. Long-read sequencing of diagnosis and post-therapy medulloblastoma reveals complex rearrangement patterns and epigenetic signatures. *Cell Genom* **2023**; 3:100281.
79. Ben Khedher M, Ghedira K, Rolain J-M, Ruimy R, Croce O. Application and challenge of 3rd generation sequencing for clinical bacterial studies. *Int J Mol Sci* **2022**; 23:1395.
80. Khezri A, Avershina E, Ahmad R. Hybrid assembly provides improved resolution of plasmids, antimicrobial resistance genes, and virulence factors in *Escherichia coli* and *Klebsiella pneumoniae* clinical isolates. *Microorganisms* **2021**; 9:2560.