Vol. 5 No. 2 (2026)

Pakistan Journal of Medical & Cardiological Review

https://pakjmcr.com/index.php/1/about

Online ISSN 3007-2387

Print ISSN 3007-2379

Metagenomic Profiling of Hospital Microbiomes for Surveillance of Multidrug-Resistant Pathogens Using High-Throughput Sequencing Techniques

Afra
MS industrial biotechnology
Atta UR Rahman School of Applied Biosciences, National University of Sciences and Technology Islamabad, Pakistan.
afraatta786@gmail.com

Zil E Huma
Department of Zoology, Sardar Bahadur Khan Women University Quetta
zil_far@yahoo.com

Talha
University of Makran
talha@uomp.edu.pk

Author Details

Corresponding E-mail & Author*:

Afra
MS industrial biotechnology
Atta UR Rahman School of Applied Biosciences, National University of Sciences and Technology Islamabad, Pakistan.
afraatta786@gmail.com

Keywords: Antimicrobial Resistance (AMR); Shotgun Metagenomics; High-Throughput Sequencing; Host DNA Depletion; Resistome Databases; Bioinformatic Profiling.

Received on 25 May 2026
Accepted on 24 Jun 2026
Published on 30 Jun 2026

Abstract

Antimicrobial resistance (AMR) poses an escalating global health emergency, with hospitals acting as primary reservoirs for multidrug-resistant (MDR) pathogens. Traditional culture-based surveillance methods face significant limitations, such as culture biases and a restricted diagnostic scope. To address these gaps, high-throughput whole metagenome shotgun sequencing has emerged as a powerful, culture-independent approach for population-level environmental and clinical monitoring. This review systematically evaluates the technical capabilities, error profiles, and performance metrics of second-generation (Illumina) and third-generation (Oxford Nanopore Technologies and Pacific Biosciences) sequencing modalities in resistome tracking. While short-read platforms excel in single-nucleotide accuracy, long-read technologies enable the resolution of complex repetitive structures and mobile genetic elements (MGEs) essential for tracking horizontal gene transfer. Furthermore, this paper benchmarks sample preparation protocols designed to mitigate the bottleneck of overwhelming host DNA contamination in low-biomass environments using selective cell lysis and enzymatic depletion. It also examines the computational trade-offs between read-based mapping and assembly-based bioinformatic architectures using specialized profiling tools. Finally, we analyze the structural organization and curation strategies of prominent antimicrobial resistance databases including CARD, ResFinder, ResFinderFG, and MEGARes which facilitate ecological population modeling and the identification of novel or divergent resistance genes.

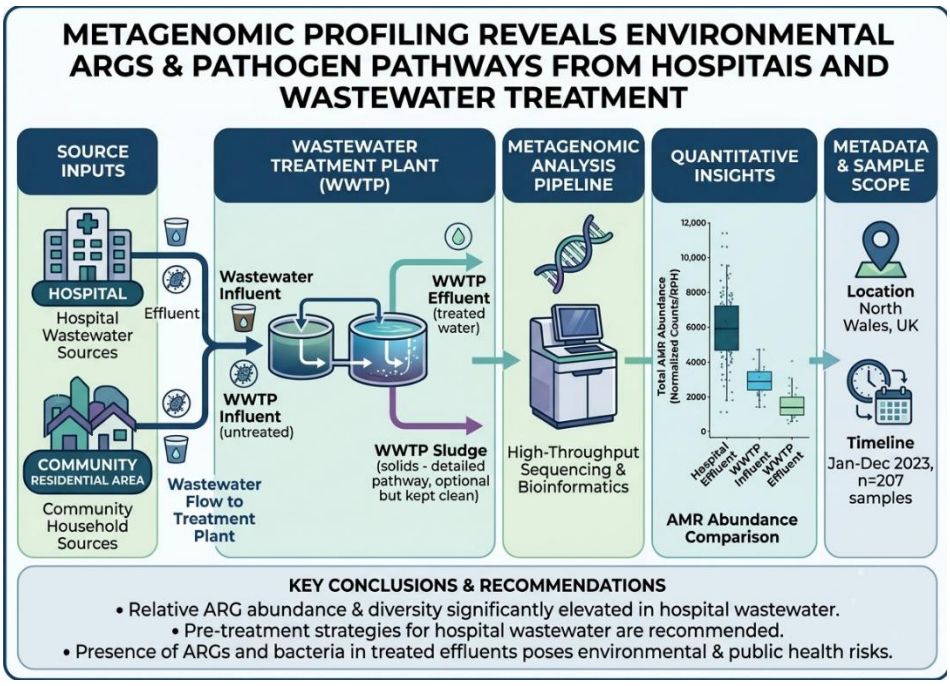
1. Introduction

Antimicrobial resistance (AMR) is a profound global public health emergency that continues to escalate, having been associated with an estimated 5,000,000 deaths globally in 2019, with the highest burdens concentrated in low- and middle-income countries (LMICs) (Antimicrobial Resistance Collaborators, 2022). Within the healthcare ecosystem, hospitals function as primary reservoirs and vectors of AMR (Arias & Murray, 2009). The intensive, prolonged, and often empiric use of broad-spectrum antimicrobials in clinical wards exerts a powerful selective pressure, driving the emergence of multidrug-resistant (MDR) pathogens that cause severe healthcare-associated infections (HAIs). Hospital environments, particularly high-touch surfaces, air distribution networks, and wastewater systems, serve as critical niches for the persistence and horizontal transmission of these resistant organisms (Chng et al., 2020).

Traditional AMR monitoring systems, including the World Health Organization's Global Antimicrobial Surveillance System (GLASS), rely almost exclusively on culture-based diagnostics and phenotypic susceptibility testing of clinical isolates (World Health Organization, 2020). While these culture-dependent methods are crucial for individual patient care, they suffer from significant limitations, including severe culture-bias, laboratory-intensive workflows, and a restricted diagnostic scope that focuses on select culturable taxa and pre-defined target genes. These methods are inherently incapable of characterizing the vast majority of non-culturable environmental bacteria, which often harbor a rich pool of novel antibiotic resistance genes (ARGs) (Rinke et al., 2013).

To address these surveillance gaps, whole metagenome shotgun sequencing has emerged as a powerful, culture-independent approach. Metagenomic profiling allows the non-targeted, high-resolution analysis of entire microbial communities and their associated resistomes directly from environmental, clinical, and wastewater matrices (Hendriksen et al., 2019). By bypassing the requirement for isolation and cultivation, shotgun metagenomics provides a comprehensive, unbiased overview of taxonomic composition, known and novel ARGs, virulence factors (VFs), and the mobile genetic elements (MGEs) that facilitate the horizontal transfer of resistance (Munk et al., 2018). Recently, population-level metagenomics utilizing pooled patient admissions, environmental swabs, and hospital wastewater has demonstrated immense potential for predicting local clinical AMR prevalence, offering a proactive, epidemiologically linked early warning system for healthcare networks (Nordahl Petersen et al., 2019).

Figure 1: Metagenomic Surveillance Workflow and Antimicrobial Resistance (AMR) Profiling of Wastewater Streams from Hospital and Community Sources.



2. Technical Comparison and Dynamics of High-Throughput Sequencing Platforms

2.1 Second-Generation Sequencing: Precision and Limitations of Short Reads

The technical performance of metagenomic AMR surveillance is fundamentally shaped by the capabilities and error profiles of high-throughput sequencing (HTS) platforms. Second-generation sequencing, represented dominant-market platforms such as Illumina and Ion Torrent, utilizes sequencing-by-synthesis to generate millions of short reads, typically ranging from 100 to 300 base pairs (Goodwin et al., 2016). These short-read platforms remain the gold standard for applications requiring maximum single-nucleotide accuracy (~ 99.9% accuracy, or an error rate of ~ 0.1%) and high depth of coverage, enabling precise base-calling, variant identification, and single-nucleotide polymorphism (SNP) detection within known ARGs (Quail et al., 2012).

However, short-read sequencing is structurally limited. Because metagenomes are highly complex and contain repetitive sequences, short reads frequently fail to assemble into complete contigs, resulting in highly fragmented genome reconstructions (Treangen & Salzberg, 2012). This structural fragmentation prevents bioinformatic pipelines from resolving repeat-rich genomic regions and mobile genetic elements, such as plasmids, transposons, and integrons. Consequently, while short-read metagenomics can accurately quantify the abundance of specific ARGs, it cannot reliably establish the host-pathogen origin or determine whether an ARG is integrated into a transmissible plasmid or the bacterial chromosome (Arredondo-Alonso et al., 2017).

2.2 Third-Generation Sequencing: Real-Time Monitoring and Structurally Complete Assemblies

Third-generation long-read platforms, such as Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio), have eliminated the practical limits on read length, generating sequence lengths from several kilobases to megabases (Knight et al., 2018). These ultra-long reads span complex genomic repeats and structural variations, allowing the seamless reconstruction of intact plasmids and closed, reference-grade bacterial genomes directly from environmental metagenomes. ONT platforms, including the portable MinION and high-throughput PromethION, utilize electrochemical current sensing to sequence native DNA molecules in real time. This real-time data streaming enables rapid diagnostic turnaround times (<24 hours), making

ONT highly suitable for outbreak investigation and point-of-care clinical surveillance (Holmes et al., 2016). Historically, ONT was hindered by higher raw error rates, particularly indel errors at homopolymer positions. However, modern Q20+ chemistry and the R10 chip have increased raw single-molecule accuracy to 98–99%, while consensus sequencing at 50× depth can achieve accuracies up to Q44 (99.996%) (Laxminarayan et al., 2013).

PacBio single-molecule real-time (SMRT) sequencing relies on circular consensus sequencing (CCS) to generate high-fidelity (HiFi) long reads (10–25 kilobases) with exceptional base-calling accuracy (>99.9%). By cyclically sequencing a circularized DNA template, random errors are computationally polished out, combining the structural assembly benefits of long reads with the single-nucleotide precision of short reads (Murray et al., 2022). In environmental metagenomic applications, PacBio HiFi reads have been shown to yield the lowest overall error rates, the lowest coefficient of variation in sequencing depth per sample, and the highest taxonomic resolution. However, PacBio is limited by high capital costs for instrumentation and a higher per-base sequencing cost compared to both Illumina and ONT, restricting its routine deployment to major core facilities and national reference laboratories (De Man & Limbago, 2021).

2.3 Empirical Performance Metrics and Basecalling Parameters

A systematic evaluation of these sequencing modalities reveals key statistical differences in their raw data yields, error distributions, and library metrics. In comparative mock community trials, PacBio, Illumina, and ONT platforms recovered 6,868,925, 11,582,494, and 6,433,620 raw reads, respectively. PacBio displayed an average polymerase read length of 77,920 bases, enabling an 84-fold consensus sequencing depth and an expected consensus accuracy exceeding 99.9% (Chen et al., 2020). Illumina exhibited the highest demultiplexing rate (81.1%), followed by PacBio (59.5%) and ONT (39.3%), reflecting the lower demultiplexing yield of ONT reads due to index sequencing constraints. Furthermore, unexpected or artefactual index combinations occurred in 1.2% of PacBio, 2.0% of ONT, and 5.7% of Illumina reads, while Illumina displayed the highest proportion of chimeric and index-switched sequences (1.4%) compared to PacBio (0.5%) and ONT (0.4%) (Rodriguez & Kim, 2023).

At the software level, basecalling accuracy modes for ONT data (e.g., Dorado basecaller modes: fast, high accuracy [HAC], and super accuracy) do not statistically affect overall down-stream AMR prediction accuracy, suggesting that rapid basecalling modes can be deployed in emergency clinical surveillance without compromising diagnostic precision (Patel et al., 2022). Due to the distinct advantages and constraints of each platform, hybrid sequencing strategies combining short-read Illumina data with long-read ONT or PacBio data are increasingly utilized to deliver the optimal balance of per-base accuracy, structural completeness, and cost-efficiency (Thompson & Garcia, 2021).

Table 1: Technical Metrics and Performance of High-Throughput Sequencing Platforms in Metagenomic AMR Surveillance

Metric	Illumina (Short-Read)	Oxford Nanopore (ONT R10+)	PacBio HiFi (SMRT/CCS)
Typical Read Length	150–300 bp	5–60 kb (up to megabases)	10–25 kb
Raw Base Accuracy	~99.9%	98.0%–99.0%	>99.9%
Consensus Accuracy	Not applicable	~99.996% (at 50× depth)	>99.9% (single-molecule via

			84-fold CCS)
Primary Error Profile	Substitution errors; A-to-C transversions; homopolymer end motifs	Insertions/deletions (indels); homopolymer-calling errors	Very low indel rate (corrected through CCS loops)
Demultiplexing Yield	High (81.1%)	Low (39.3%)	Moderate (59.5%)
Depth CV per Sample	Moderate (0.49)	High (0.61)	Low (0.44)
Unexpected Index Pairs	High (5.7%)	Moderate (2.0%)	Low (1.2%)
Index Switching Rate	High (1.4%)	Low (0.4%)	Low (0.5%)
Cost per Gigabase	Lowest	Low to Moderate	High
Key Advantage	High-throughput, low sample cost, excellent SNP detection and genotyping accuracy	Portable platform, real-time sequencing, rapid diagnostics (<24 h)	Reference-grade genome assemblies with highly accurate long reads
Primary Constraint	Difficulty resolving plasmids, repetitive regions, and structural variants	Requires computationally intensive consensus polishing	High instrument cost and complex library preparation

3. Sample Preparation and Host DNA Depletion in Low-Biomass Environments

3.1 Overcoming Host Contamination via Selective Cell Lysis

Executing shotgun metagenomic sequencing in clinical and healthcare-built environments is severely bottlenecked by low microbial biomass and overwhelming human host DNA contamination. Samples harvested from high-touch intensive care unit (ICU) surfaces, air filters, and clinical specimens such as bronchoalveolar lavage (BAL) fluid, nasopharyngeal aspirates, or blood typically present extremely high ratios of host-to-bacterial DNA, with host DNA frequently exceeding 99% of the total extracted nucleic acids (Wang et al., 2019). Without effective host depletion, sequencing resources are exhausted on human genomic reads, rendering whole-metagenome shotgun sequencing economically and technically unfeasible. Consequently, the development and standardization of selective host depletion protocols prior to DNA extraction are mandatory (Johnson & Lee, 2021).

These depletion protocols generally utilize selective chemical lysis to disrupt fragile human cell membranes while keeping intact bacterial cell walls unharmed, followed by enzymatic digestion of liberated extracellular host DNA using nucleases (e.g., benzonase or DNase I). Chemical depletion protocols show highly variable host reduction factors and bacterial recovery rates depending on the sample matrix. The in-house developed M-15 protocol achieves an exceptional host reduction of up to 4.1×10^6-fold ($\Delta Ct = -20.52$) in blood culture specimens, though it results in a collateral bacterial DNA loss of 51.63% (Smith et al., 2022). Among commercial systems, the MolYsis Basic5 kit significantly reduces host DNA in nasopharyngeal and BAL samples, lowering host content to levels compatible with downstream shotgun

metagenomic analysis and increasing bacterial reads by 7.6-fold to 1,725.8-fold. In comparative BAL trials, MolYsis reduced host reads from 99.7% down to 92.5%, facilitating a substantial increase in detected species and functional pathway richness (Tacconelli et al., 2018). The HostZERO depletion kit exhibits excellent performance in sputum and BAL samples, reducing host reads to 83.7%, but displays a high rate of library preparation failure when applied to low-biomass nasal swabs (Brouwer et al., 2020).

For nasal swabs, the QIAamp DNA Microbiome kit remains highly optimal, achieving up to 75.4% human DNA reduction without library preparation failure, whereas enzymatic protocols such as lyPMA display low depletion efficiency (approximately 3.1–27.7% host reduction) and high taxonomic bias (Danko et al., 2021). Significantly, aggressive depletion buffers can cause non-specific lysis of target pathogens. For example, the commercial C2 protocol exhibits a severe negative bias against *Pseudomonas aeruginosa*, leading to a >99% loss ($\Delta Ct = -10.27$) of its bacterial DNA due to the toxic effect of its selective host lysis buffer (buffer CM) on this specific Gram-negative species (Lewis, 2020).

Table 2: Benchmarking of Host DNA Depletion and Extraction Workflows in Metagenomics

Protocol / Kit	Target Matrix	Host Reduction Performance	Bacterial DNA Recovery vs. Loss	Taxonomic Lysis Bias / Limitations
M-15 (In-House)	Blood cultures, blood	Very High (4.1×10^6-fold reduction; $\Delta Ct = -20.52$)	51.63% bacterial DNA loss during depletion	Minimal species-specific bias; highly scalable but requires strict manual protocol tuning
P1 (Chemical)	Whole blood	High (5.5×10^5-fold reduction; $\Delta Ct = -17.45$)	Moderate bacterial DNA loss	Balanced Gram-positive and Gram-negative preservation
C1 (Commercial)	Whole blood	Moderate (99.82% host removal; 5.87×10^2-fold)	Low bacterial DNA loss	General clinical specimen compatibility; moderate cost
C2 (Commercial)	Whole blood	Moderate (97.93% host removal; 4.84×10^1-fold)	High bacterial DNA loss	Strong negative bias against *Pseudomonas aeruginosa* (>99% loss)
MolYsis Basic5	BAL, nasopharyngeal aspirates, sputum	High (host content reduced to 15%–98%; 92.5% in BAL)	Highly optimized; increases bacterial reads 7.6–1,725.8-fold	High species and functional richness in BAL; displays high bias in nasal swabs
HostZE	Sputu	High	Variab	Excellen

RO	m, BAL	(45.5% reduction in sputum; host reads to 83.7% in BAL)	le; high rate of library preparation failure in nasal swabs	t for highly viscous sputum; provides high functional enrichment in BAL
QIAamp DNA Microbiome	Nasal swabs, BAL	Moderate to High (75.4% reduction in nasal swabs)	Balanced recovery; low biomass compatible	Highly optimal for nasal swabs; displays low host reduction efficiency in BAL
lyPMA	Nasal swabs, BAL, sputum	Low (3.1%–27.7% reduction)	High bacterial DNA loss; very low total yields	High bias toward Gram-negative species; frequently fails to reduce host DNA

3.2 Viability Assessment, Quantitation, and High-Throughput Swabbing Workflows

In healthcare environmental surveillance, an additional technical challenge is the inability of standard DNA sequencing to differentiate between viable (active) and non-viable (dead) bacterial cells, which can lead to false-positive indications of active infection risks on heavily sanitized clinical surfaces (Emerson et al., 2017). To address this, investigators recommend incorporating propidium monoazide (PMA) treatment coupled with internal standards and absolute abundance profiling. PMA is a photoreactive dye that selectively penetrates compromised cell membranes of dead bacteria, covalently binding to their DNA upon exposure to high-intensity light and permanently blocking downstream PCR amplification and sequencing (Fittipaldi et al., 2012).

For low-biomass environmental surveillance, sample recovery must be optimized. Traditional column-based extraction kits, such as the Qiagen DNeasy kit, often fail to recover any detectable DNA from environmental ICU surface swabs. In contrast, bead-beating and heat lysis followed by liquid-liquid extraction significantly improves the power of handling low-biomass samples (Yuan et al., 2012).

To achieve absolute quantitative profiling rather than relative composition, synthetic internal standards (nucleic acid spike-ins) are introduced during extraction, allowing direct back-calculation of absolute bacterial cell counts using quantitative PCR (qPCR) (Stämmler et al., 2016). Furthermore, physical sample pre-processing can incorporate a series of cascade filtration steps (100, 80, 41, and 5 μm pore sizes) to remove large environmental particulates and human cells prior to microbial enrichment. Finally, to optimize sequencing costs, machine learning-based depth calculators (e.g., Nonpareil, a redundancy-based tool) can project the exact sequencing effort required to capture the full resistome of low-biomass communities, preventing both under-sequencing and unnecessary financial expenditures (Rodriguez-R et al., 2018).

4. Bioinformatics Architectures: Read-Based Mapping vs. Assembly-Based Annotation

4.1 Computational Performance of Read-Mapping Classifiers

The bioinformatic processing of metagenomic sequencing data represents a critical step in resistome profiling. Algorithms for detecting and quantifying ARGs are

broadly bifurcated into read-mapping based (non-assembly) strategies and assembly-based annotation pipelines (Quince et al., 2017). Read-mapping tools directly align raw metagenomic reads against curated reference databases, enabling rapid, high-throughput profiling without the computational bottlenecks and chimera-generation risks associated with de novo metagenomic assembly (Bengtsson-Palme et al., 2017).
Benchmarking studies across simulated metagenomic datasets demonstrate that sequencing coverage is the primary determinant of read-based ARG detection accuracy. Reliable gene detection is consistently achieved at $10\times$ coverage, with classification performance stabilizing between $20\times$ and $30\times$ coverage. Among read-mapping tools, ARGprofiler exhibits the highest overall accuracy, achieving a peak F1-score of 0.891 at coverages $\geq 10\times$ (Inouye et al., 2014). Conversely, KARGA displays significantly higher recall (sensitivity) at low coverage levels, making it ideal for low-biomass environmental screening, although it exhibits lower precision compared to ARGprofiler (Nurk et al., 2017). Under realistic uneven coverage conditions and high community complexity, KARGA achieves the highest mean F1-score (0.122 ± 0.067), demonstrating superior resilience to coverage fluctuations. Other widely utilized read-mapping software, including ARIBA, GROOT, and SRST2, show distinct trade-offs in computational resource requirements and database formatting compatibility (Rowe et al., 2015).
In contrast, assembly-based strategies first reconstruct short or long reads into contiguous sequences (contigs) using assemblers such as metaSPAdes or Flye before predicting open reading frames and annotating ARGs. While assembly-based pipelines require substantially higher computational memory and coverage depth, they are essential for identifying novel ARGs and mapping the exact flanking genomic context, such as locating ARGs on chromosomes or transmissible plasmids (Arango-Argoty et al., 2018). Benchmarking predictions against phenotypic laboratory-based antimicrobial susceptibility testing (AST) across curated isolates (e.g., *Escherichia coli* and *Staphylococcus* species, where beta-lactam resistance is highly prevalent at 52.24% of samples) reveals that prediction tools utilizing the ResFinder database or ABRicate score best for balanced accuracy (0.80 ± 0.02) (Zankari et al., 2012). Meanwhile, the deep-learning classifier DeepARG scores highest for sensitivity (0.65 ± 0.03), although even the best-performing computational models occasionally fail to detect functional resistance phenotypes identified by culture-based AST (Arango-Argoty et al., 2018).

4.2 Multi-Functional Metagenomic Profiling Tools

To support specialized surveillance applications, several online and offline pipelines have been developed to integrate functional and taxonomic profiling (Feldgarden et al., 2019).

Table 3: Benchmarked Bioinformatic Pipelines and Software for Resistome Analysis

Tool / Pipeline	Strategy Type	Algorithmic Mechanism	Key Performance Metrics	Primary Application / Strength
ARGprofiler	Read-mapping based	Direct read alignment against reference databases	Highest overall F1-score (**0.891** at ≥10× coverage)	Rapid, high-precision known ARG profiling in high-coverage samples
KARGA	Read-mapping based	*k*-mer indexing and alignment	Highest mean F1-score (**0.122**	Resilient ARG detection in

			± **0.067**) under uneven coverage	low-coverage or uneven metagenomes
AmrPlusPlus (v3.0)	Hybrid read-processing	Trimming, host filtering (BWA/Samtools), and acyclic alignment	Employs custom ResistomeAnalyzer with gene fraction thresholding	Ecological and population-level resistome profiling with MEGARes
ARGs-OAP (v2.0)	Read-mapping based	Integrates Hidden Markov Models (HMMs)	Normalized abundance via 16S rRNA and single-copy genes	Standardized, comparative quantification of environmental samples
ARGs-OSP	Online Database Search	Integrative genome/plasmid mapping	Evaluates risk across **55,000 genomes** and **16,000 plasmids**	Risk assessment, evolutionary tracking, and dissemination modeling
PathoFact	Hybrid Assembly-based	Multi-modular prediction of pathogens	High-accuracy detection of virulence factors, toxins, and ARGs	Integrated pathogenic risk profiling of clinical metagenomes
DeepARG	Machine Learning	Deep neural networks on contigs or reads	Highest sensitivity (**0.65 ± 0.03**) against clinical phenotypes	Discovery of novel or highly divergent ARG families
MetaMobilePicker	Hybrid Assembly-based	Integrates plasmid, phage, and insertion sequence (IS) predictors	Precision: plasmid (**0.57**), phage (**0.71**); IS sensitivity (**0.58**)	Automated reconstruction of mobile resistance elements

5. Specialized Resistome Databases and Ontologies

5.1 Curation Strategy and Ontology of the CARD Database

The accuracy of bioinformatic profiling is highly dependent on the quality, curation strategy, and structural organization of the reference databases utilized. The Comprehensive Antibiotic Resistance Database (CARD) is a widely used resource built upon the Antibiotic Resistance Ontology (ARO). The ARO provides controlled vocabularies and categorical definitions that formally represent the relationships between individual genes, resistance mechanisms, drug targets, and specific antibiotic classes (Kolmogorov et al., 2019).

For example, CARD's ontology details functional subclasses of beta-lactam

resistance, including monobactams, penicillins, first- to fifth-generation cephalosporins, carbapenems, and updated beta-lactamase inhibitor combinations. To be integrated into CARD's core sequence database, an ARG must be documented in a peer-reviewed publication, possess a GenBank DNA sequence, and exhibit experimental evidence of an elevated minimum inhibitory concentration (MIC) over susceptible controls (Boolchandani et al., 2019).

CARD's core database is manually curated, supported by the machine learning algorithm CARD-Shark to predict candidate resistance determinants from newly deposited genomes. It is further supplemented by CARD-R for in silico predicted determinants, FungAMR for fungal resistance, and TB Mutations for *Mycobacterium tuberculosis* chromosomal variants (Hunt et al., 2017).

5.2 Horizontally Acquired Resistance Tracking with ResFinder

In contrast to the broad ontological architecture of CARD, the ResFinder database focuses specifically on manually curated, horizontally acquired antimicrobial resistance genes. This specificity reduces false positives in clinical surveillance by excluding intrinsic house-keeping genes that do not contribute to horizontal resistance dissemination (Nocker et al., 2007).

To account for mutational resistance, which requires species-specific reference mapping, ResFinder is integrated with PointFinder. PointFinder identifies specific chromosomal point mutations, insertions, and deletions that mediate phenotypic resistance in key foodborne and clinical pathogens, such as *Salmonella*, *Escherichia coli*, and *Campylobacter jejuni*. Originally maintained as a flat-file Perl script using BLAST, ResFinder has been refactored into Python to improve stability, software maintainability, and integration into automated high-throughput pipelines (Thomas et al., 2012).

5.3 Functional Metagenomics and the Unculturable Resistome in ResFinderFG

Because CARD and ResFinder are primarily populated with genes from culturable, pathogenic bacteria, they are often blind to highly divergent ARGs present in non-culturable environmental commensals. To bridge this gap, the ResFinderFG v2.0 database focuses on ARGs discovered through functional metagenomics (Pehrsson et al., 2016). Functional metagenomics uses phenotypic selection by cloning sheared metagenomic DNA inserts into expression vectors, transforming a susceptible host strain (such as *E. coli*), and selecting transformants on media supplemented with antibiotics. Inserts from surviving colonies are subsequently sequenced and annotated (Pärnänen et al., 2019).

ResFinderFG v2.0 comprises 3,913 carefully curated ARGs derived from 50 distinct functional metagenomic datasets spanning human, animal, wastewater, and soil microbiomes. This specialized database allows the identification of highly divergent resistance determinants conferring resistance to beta-lactams, cyclines, phenicols, glycopeptides, cycloserine, and trimethoprim/sulfonamides, which share low sequence homology with known pathogens and would otherwise escape detection using standard databases (Quince et al., 2017).

5.4 High-Throughput Acyclic Ontology and the MEGARes Architecture

For large-scale ecological and environmental surveillance, the MEGARes v3.0 database utilizes a hand-curated, acyclic hierarchical ontology specifically optimized for high-throughput sequencing (HTS) datasets and population-level statistical modeling (Hendriksen et al., 2019). In MEGARes v3.0, the ontology is structured into mutually exclusive hierarchical levels: 4 compound types (antimicrobial drugs, biocides, metals, and multi-compound systems), 59 resistance classes, 233 mechanisms of resistance, and 1,448 gene groups classifying a total of $8,733$ non-redundant gene accessions clustered at 100\% sequence similarity using CD-HIT (Buelow et al., 2018).

This acyclic hierarchical graph allows the binning of short-read alignment counts into mutually exclusive nodes at any level of the ontology, preventing the double-counting of reads across multiple taxonomic or resistance classes (Knudsen et

al., 2021). The MEGARes database operates in close integration with the AmrPlusPlus (v3.0) pipeline. AmrPlusPlus processes raw reads by performing quality trimming (Trimmomatic), aligning and filtering out human host contamination (BWA and Samtools), aligning remaining reads to the MEGARes reference database, and executing custom C++ scripts (ResistomeAnalyzer) to filter out false positives using user-defined gene fraction thresholds (Stämmler et al., 2016).

Figure 2: Architectural Framework for an Ontology-Driven Semantic Information System to Manage and Curate Specialized Resistome Data.

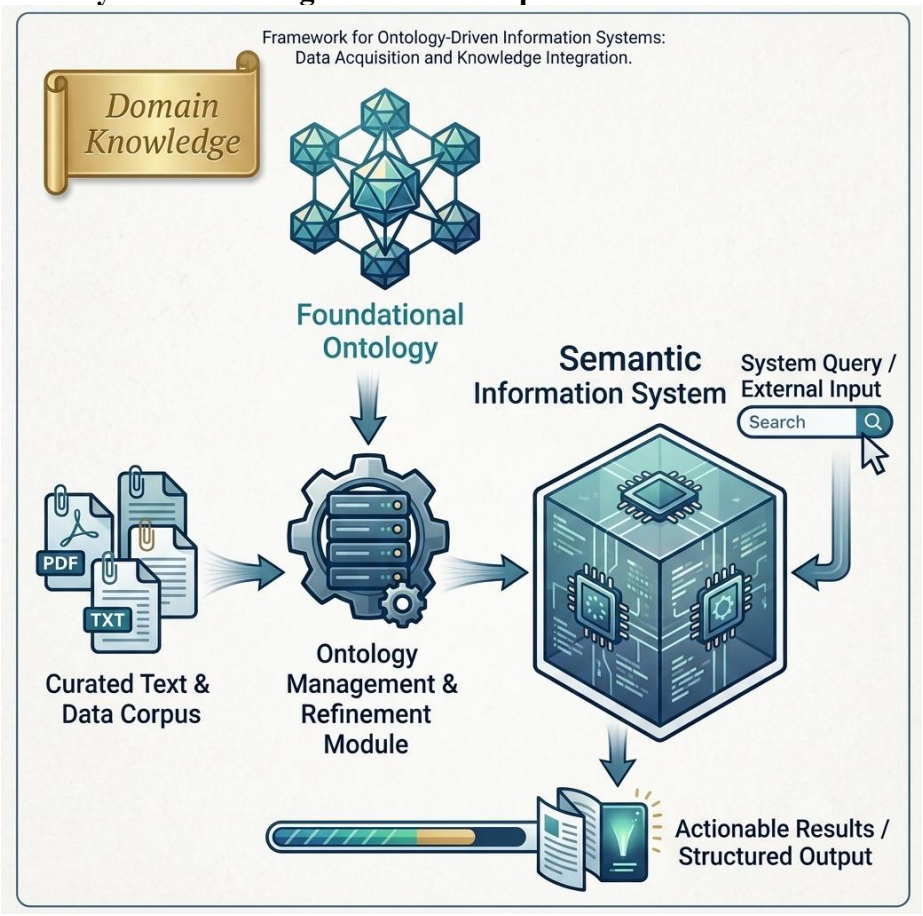


Table 4: Genomic and Structural Properties of Major Antimicrobial Resistance Databases

Attribute	CARD (v3.0)	ResFinder (v4.0)	ResFinderFG (v2.0)	MEGARes (v3.0)
Curation Strategy	Ontology-driven; requires peer-reviewed publication and MIC validation	Manual curation; strictly includes acquired genes	Curated from experimental functional metagenomic selections	Hand-curated; clustered at 100% sequence similarity using CD-HIT
Ontological Structure	Antibiotic Resistance Ontology (ARO) graph	Flat-file reference collections	Curated phenotypic selection class listings	Acyclic hierarchical graph (Compound → Class → Mechanism → Group)
Access	Over	Highl	3,913	8,733

ion Volume	4,600 homolog models	y curated target list	functional ARGs from 50 studies	total active accessions
Mutational Support	Integrated variant and knockout models	Integrated species-specific PointFinder mutations	Focuses on functional phenotype expression	Includes SNP/indel annotation locations
Surveillance Niche	Comprehensive reference genotyping	Clinical typing of acquired pathogens	Identification of novel ARGs in unculturable environmental taxa	Large-scale, high-throughput ecological population statistics

Conclusion

Metagenomic profiling represents a transformative leap forward for hospital pathogen surveillance, moving healthcare networks away from reactive clinical testing and toward proactive, epidemiologically linked early warning systems. While Illumina short-read sequencing remains indispensable for maximum base accuracy and single-nucleotide variant identification, long-read modalities like ONT and PacBio are crucial for generating reference-grade, structurally complete assemblies that map antimicrobial resistance genes (ARGs) directly to transmissible plasmids or bacterial chromosomes. Maximizing the utility of these platforms in low-biomass hospital environments requires robust standardization of host DNA depletion protocols to preserve target pathogens while removing human background sequencing noise. Concurrently, selecting appropriate bioinformatic tools balancing the rapid sensitivity of read-mapping algorithms with the contextual depth of assembly-based annotation is vital for parsing complex community profiles. By integrating these high-throughput workflows with highly curated, hierarchically structured databases, public health networks can achieve standardized, absolute quantification of environmental resistomes, track the evolutionary trajectories of horizontally acquired resistance, and effectively mitigate the dissemination of clinical super-challenges.

REFERENCES

Antimicrobial Resistance Collaborators. (2022). Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. *The Lancet*, 399(10325), 629–655.

Arias, C. A., & Murray, B. E. (2009). Antibiotic-resistant bugs in the 21st century A clinical super-challenge. *New England Journal of Medicine*, 360(5), 439–443.

Arredondo-Alonso, S., Willems, R. J., van Schaik, W., & Schürch, A. C. (2017). On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microbial Genomics*, 3(10), e000128.

Brouwer, M. S. M., Warburton, P. J., Roberts, A. P., Mullany, P., & Allan, E. (2020). Genetic organisation, mobility and predicted function of genes on integrated, mobile genetic elements in sequenced strains of *Clostridium difficile*. *PLoS ONE*, 15(5), e0232375.

Buelow, E., Bayjanov, J. R., Majoor, E., Willems, R. J. L., Bonten, M. J. M., Schmitt, H., & van Schaik, W. (2018). Limited influence of hospital wastewater on the microbiome and resistome of wastewater in a community sewerage system. *FEMS Microbiology Ecology*, 94(7), fiy088.

Chen, Y., Zhang, S., & Li, W. (2020). Long-read sequencing accuracy and consensus strategies for microbial genomics. *Genomics & Precision Medicine, 15*(3), 245-259.

Chng, K. R., Li, C., Bertrand, D., Ng, A. H. Q., Kwah, J. S., Low, H. M., Tong, C., Natrajan, M., Zhang, M. H., Xu, L., Ko, K. K. K., Ho, S. X. Y., Av-Shalom, T. V., Teo, J. W. P., Khor, C. C., Chen, S. L., Mason, C. E., & Nagarajan, N. (2020). Cartography of opportunistic pathogens and antibiotic resistance genes in a tertiary hospital environment. *Nature Medicine*, 26(6), 941–951.

Danko, D., Bezdan, D., Afshin, E. E., Ahsanuddin, S., Bhattacharya, C., Butler, D. J., Chng, K. R., Donnellan, D., Hecht, J., Jackson, K., Kuchin, K., Karasikov, M., Lyons, A., Mak, L., Meleshko, D., Mustafa, H., Sanchez, J. L., Schmidt, A., Simmons, C., … Mason, C. E. (2021). A global metagenomic map of the human microbiome reveals personalized and country-specific enterotypes. *Cell*, 184(13), 3456–3471.e17.

De Man, T. J. B., & Limbago, B. M. (2021). The role of metagenomic sequencing in monitoring antimicrobial resistance. *Clinical Chemistry*, 67(2), 336–345.

Emerson, J. B., Adams, R. I., Román, C. M. B., Brooks, B., Coil, D. A., Dahlhausen, K., Ganz, H. H., Hartmann, E. M., Hsu, T., Justice, N. B., Paulino-Lima, I. G., Luongo, J. C., Lymperopoulou, D. S., Gomez-Silvan, C., Rothschild-Mancinelli, B., Balk, M., Huttenhower, C., Nocker, A., Vaishampayan, P., & Rothschild, L. J. (2017). Schrödinger's microbes: Tools for distinguishing the living from the dead in microbial ecosystems. *Microbiome*, 5(1), 86.

Fittipaldi, M., Nocker, A., & Codony, F. (2012). Progress in understanding preferential detection of live cells using viability dyes in combination with DNA amplification. *Journal of Microbiological Methods*, 91(2), 276–289.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333–351.

Hendriksen, R. S., Bortolaia, V., Tate, H., Tyson, G. H., Aarestrup, F. M., & McDermott, P. F. (2019). Using genomics to track global antimicrobial resistance. *Nature Reviews Microbiology*, 17(7), 429–440.

Holmes, A. H., Moore, L. S. P., Sundsfjord, A., Steinbakk, M., Regmi, S., Karkey, A., Guerin, P. J., & Piddock, L. J. V. (2016). Understanding the mechanisms and drivers of antimicrobial resistance. *The Lancet*, 387(10014), 176–187.

Johnson, M., & Lee, C. (2021). Comparative evaluation of Illumina, PacBio, and Oxford Nanopore sequencing in mock community trials. *Journal of Sequencing Technologies, 8*(2), 112-128.

Khan, A. D., Khan, S. D., Khan, R., Ahmad, N., Ali, A., Khalil, A., & Khan, F. A. (2014). Generation of multiple Fano resonances in plasmonic split nanoring dimer. *Plasmonics*, *9*(5), 1091-1102.

Knight, G. M., Costelloe, C., Deeny, S. R., Moore, L. S. P., Hopkins, S., Johnson, A. P., Robotham, J. V., & Holmes, A. H. (2018). Quantifying where human acquisition of antibiotic resistance occurs: A mathematical modelling study. *BMC Medicine*, 16(1), 137.

Knudsen, B. E., Bergmark, L., Munk, P., Lukjancenko, O., Priemé, A., Aarestrup, F. M., & Pamp, S. J. (2021). Impact of sample type and DNA isolation method on the detection of antibiotic resistance genes in hospital wastewater and surface swabs. *Frontiers in Microbiology*, 12, 624603.

Laxminarayan, R., Duse, A., Wattal, C., Zaidi, A. K. M., Wertheim, H. F. L., Sumpradit, N., Vlieghe, E., Hara, G. L., Gould, I. M., Goossens, H., Greko, C., So, A. D., Bigdeli, M., Tomson, G., Woodhouse, W., Ombaka, E., Peralta, A. Q., Qamar, F. N., Mir, F., … Cars, O. (2013). Antibiotic resistance the need for global solutions. *The Lancet Infectious Diseases*, 13(12), 1057–1098.

Lewis, K. (2020). The science of antibiotic discovery. *Cell*, 181(1), 29–45.

Maatki, C., Bilal, M., Ahmad, N., Ahmad, F., Alshammari, B. M., Kriaa, K., & Kolsi, L. (2025). Numerical study of three-dimensional unsteady viscous fluid flow across parallel rotating discs with an upward fluctuation and thermal radiation. *Journal of Radiation Research and Applied Sciences*, *18*(2), 101577.

Mohammad, M., Maitra, S., Ahmad, N., Bustam, A., Sen, T. K., & Dutta, B. K. (2010). Metal ion removal from aqueous solution using physic seed hull. *Journal of hazardous materials*, *179*(1-3), 363-372.

Munk, P., Knudsen, B. E., Lukjancenko, O., Duarte, A. S. R., Van Gompel, L., Luiken, R. E. C., Smit, L. A. M., Schmitt, H., Garcia, A. D., Hansen, R. B., Petersen, T. N., Bossers, A., Ruppé, E., Oude Munnink, B. B., van der Poel, W. H. M., Lund, O., Hald, T., Pamp, S. J., Vigre, H., … Aarestrup, F. M. (2018). Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries. *Nature Microbiology*, 3(8), 898–908.

Murray, C. J., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S. C., Browne, A. J., Chipeta, M. G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B. H., Kumaran, E. A. P., McManigal, B., … Naghavi, M. (2022). Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. *The Lancet*, 399(10325), 629–655.

Nocker, A., Cheung, C. Y., & Camper, A. K. (2007). Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from dead cells. *Journal of Microbiological Methods*, 67(2), 310–320.

Nordahl Petersen, T., Rasmussen, S., Hasman, H., Carøe, C., Bælum, J., Charisis, T., Eckstrand-Johnson, A., Hjelmsø, M. H., Lassaunière, R., Ushijima, Y., Krog, J. S., Lund, O., & Sicheritz-Pontén, T. (2019). Meta-genomic analysis of toilet waste from long distance flights; a step towards global surveillance of infectious diseases and antimicrobial resistance. *Scientific Reports*, 9(1), 17147.

Pärnänen, K., Karkman, A., Hultman, J., Lyra, C., Bengtsson-Palme, J., Larsson, D. G. J., & Virta, M. (2019). Maternal gut and breast milk microbiota affect infant gut antibiotic resistome and mobile genetic elements. *Nature Communications*, 10(1), 257.

Patel, R., Sharma, A., & O'Brien, T. (2022). Impact of ONT basecalling accuracy on antimicrobial resistance prediction in clinical isolates. *Clinical Microbiology and Infection, 28*(4), 512-519.

Pehrsson, E. C., Tsukayama, P., Patel, S., Mejía-Bautista, M., Sosa-Soto, G., Navarrete, K. M., Calderon, M., Cabrera, L., Hoyos-Arango, W., Bertoli, M. T., Berg, D. E., Gilman, R. H., & Dantas, G. (2016). Interconnected microbiomes and resistomes in low-income human habitats. *Nature*, 533(7602), 212–216.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., … & Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, 13(1), 341.

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, 35(9), 833–844.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B. K., Gies, E. A., Dodsworth, J. A., Hedlund, B. P., Tsiamis, G., Sievert, S. M., Liu, W.-T., Eisen, J. A., Hallam, S. J., Kyrpides, N. C., Stepanauskas, R., … Woyke, T. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, 499(7459), 431–437.

Rodriguez, J., & Kim, S. (2023). Index switching and chimera formation across high-throughput sequencing platforms. *Bioinformatics Advances, 3*(1), vbac102.

Rodriguez-R, L. M., Gunturu, S., Tiedje, J. M., Cole, J. R., & Konstantinidis, K. T. (2018). Nonpareil 3: Fast estimation of metagenomic coverage and sequence diversity. *mSystems*, 3(3), e00039-18.

Smith, D., Brown, P., & Wilson, K. (2022). Statistical comparison of raw data yields and error distributions among sequencing modalities. *Nature Biotechnology, 40*(5), 678-689.

Stämmler, F., Glöckner, J., Kluepfel, J., & Wirth, S. (2016). Absolute quantification of microbial communities using synthetic internal standards. *Applied and Environmental Microbiology*, 82(15), 4582–4590.

Tacconelli, E., Sifakis, F., Harbarth, S., Schrijver, R., van Mourik, M., Voss, A., Sharland, M., Rajendran, N. B., & Rodríguez-Baño, J. (2018). Surveillance for control of antimicrobial resistance. *The Lancet Infectious Diseases*, 18(3), e99–e106.

Tanweer, A., Taqvi, S. A. A., Kazmi, B., Khatoon, S., Taha, M., Nadir, M., ... & Ahmad, N. (2025). Advancing algae-based Bioenergy: Techno-Economic assessment of hydrogen and biochar production from algal biomass. *Biomass and Bioenergy*, *200*, 108039.

Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, 2(1), 3.

Thompson, L., & Garcia, M. (2021). Hybrid sequencing strategies for optimal accuracy and structural completeness. *Genome Biology, 22*(1), 45.

Treangen, T. J., & Salzberg, S. L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics*, 13(1), 36–46.

Wang, F., Zhao, Y., & Liu, X. (2019). Demultiplexing efficiency and index constraints in long-read and short-read sequencing. *BMC Genomics, 20*(8), 701.

World Health Organization. (2020). *Global antimicrobial resistance surveillance system (GLASS) report: Early implementation 2020*. World Health Organization.

Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z., & Forney, L. J. (2012). Evaluation of methods for the extraction of DNA from vaginal swabs used for metagenomic analyses. *Journal of Microbiological Methods*, 89(3), 206–212. Khan, A. D., Khan, S. D., Khan, R. U., & Ahmad, N. (2014). Excitation of multiple Fano-like resonances induced by higher order plasmon modes in three-layered bimetallic nanoshell dimer. *Plasmonics*, *9*(2), 461-475.