

## Review

## Open Access

# Profile surveillance and risk assessment of the environmental dimension of antibiotic resistance via the metagenomic approach

Peiju Fang<sup>1,2#</sup>, Zehui Yu<sup>1,2#</sup>, Jin Huang<sup>1,3</sup> and Bing Li<sup>1,2,4\*</sup>

Received: 9 November 2025

Revised: 8 December 2025

Accepted: 10 December 2025

Published online: 23 January 2026

## Abstract

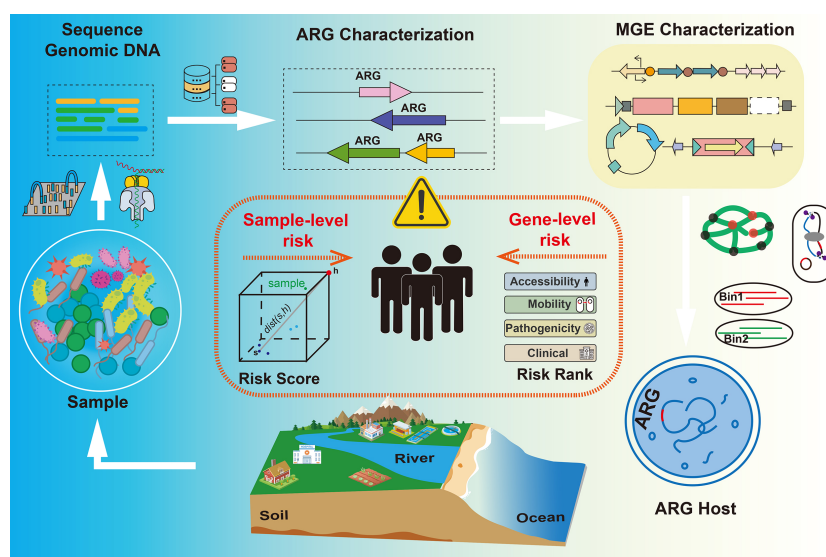
Antibiotic resistance has been recognized as a global environmental and public health challenge. Over the past decades, methods of studying antibiotic resistance genes (ARGs) have rapidly evolved. This review summarizes progress in metagenomic approaches for profiling ARGs, mobile genetic elements (MGEs), and their microbial hosts. The transition from second-generation to third-generation sequencing technologies, developments in ARG detection pipelines, and emerging strategies for absolute quantification are highlighted. In addition, novel approaches for linking ARGs to their hosts and assessing resistome risks at both gene and sample levels are discussed. Continuous improvements in methodologies are deepening the understanding of resistance dissemination, and providing a foundation for environmental surveillance and risk control.

**Keywords:** Antibiotic resistance genes, Mobile genetic elements, Metagenomics, Risk assessment, Host identification

## Highlights

- Summarized metagenomic technologies and computational tools for ARGs and MGEs profiling.
- Compared methodological advances in host identification and absolute quantification of ARGs.
- Integrated resistome risk assessment framework at gene and sample levels.

## Graphical abstract



# Authors contributed equally: Peiju Fang and Zehui Yu

\* Correspondence: Bing Li ([bingli@sz.tsinghua.edu.cn](mailto:bingli@sz.tsinghua.edu.cn))

Full list of author information is available at the end of the article.

Introduction

Antibiotic resistance has been recognized as a global threat to the environment and human health, with multidrug-resistant infections contributing significantly to the rising number of deaths worldwide<sup>[1,2]</sup>. Moreover, antibiotic resistance hinders modern medical advances, making common infections difficult to treat. Antibiotic resistance has been widely detected not only in clinical settings but also across various environments<sup>[3]</sup>. Its widespread distribution highlights the need to tackle the problem under the One Health framework<sup>[4,5]</sup>. The One Health concept emphasizes the interconnectedness of human, animal, and environmental health, recognizing that resistant bacteria and genes flow between these domains at local and global scales<sup>[4,5]</sup>. For example, antibiotic overuse in clinical and agricultural settings contributes to the release of antibiotic residues and the spread of resistant microorganisms. These have made diverse environments reservoirs of antibiotic resistance genes (ARGs), which may facilitate the transfer of resistance genes to pathogenic bacteria. Consequently, addressing antibiotic resistance requires a multidisciplinary One Health approach that integrates environmental surveillance alongside traditional clinical monitoring.

Horizontal gene transfer (HGT) plays a significant role in the rapid dissemination of ARGs<sup>[6]</sup>. In contrast to vertical gene transfer, HGT facilitates the transfer of ARGs across different bacterial species. This process is mediated by mobile genetic elements (MGEs), including plasmids, transposons, integrons, phages, and integrative and conjugative elements (ICEs), which serve as vehicles for gene transfer<sup>[7–10]</sup>. The mobility of ARGs via MGEs significantly accelerates the spread of resistance, connecting environmental and clinical resistomes. Consequently, environmental hotspots serve as breeding grounds for novel ARGs, which pathogenic bacteria may subsequently acquire via HGT<sup>[11–13]</sup>. ARGs located on conjugative plasmids or within transposable elements are generally considered to have a higher potential for dissemination than chromosomal genes<sup>[14]</sup>. As a result, current surveillance strategies increasingly incorporate the detection and characterization of plasmid-borne ARGs and integron diversity, in addition to profiling ARG abundance and diversity in environmental samples.

In response to the global antibiotic resistance crisis, advanced molecular detection technologies have become essential for identifying ARGs and assessing their associated risks<sup>[15,16]</sup>. In particular, metagenomic sequencing has greatly improved ARG surveillance by enabling culture-independent analysis of entire microbial communities<sup>[17,18]</sup>. High-throughput second-generation sequencing remains widely used for profiling ARG diversity and abundance in complex samples. Shotgun metagenomic sequencing of DNA extracted directly from samples such as water, soil, air, or feces enables the detection of ARGs in both culturable and unculturable bacteria, providing a comprehensive overview of the resistome profile<sup>[15,19]</sup>. However, the short read lengths of second-generation sequencing often impede the reconstruction of complete ARG loci or the determination of their genetic context, such as their association with MGEs or identification of hosts<sup>[20]</sup>. To overcome these

limitations, third-generation (long-read) sequencing platforms have been increasingly employed, such as Oxford Nanopore and PacBio<sup>[21,22]</sup>. These technologies can span entire ARG regions along with flanking sequences, providing information for host or plasmid inference. In parallel, advances in assembly algorithms, binning strategies, and proximity ligation methods have enhanced the resolution of host-ARG associations<sup>[23–25]</sup>. Additionally, specialized bioinformatics pipelines supported by curated reference databases have been developed to accurately identify and classify ARGs<sup>[26,27]</sup>. These tools can even detect novel or divergent resistance gene variants<sup>[28]</sup>. These developments have also enabled quantitative frameworks for assessing the public health relevance of ARGs, based on factors such as mobility, host pathogenicity, and clinical relevance<sup>[29]</sup>.

This review summarizes current sequencing technologies and bioinformatics pipelines for detecting and quantifying antibiotic resistance genes (ARGs) in metagenomic datasets. It also examines recent technical innovations and conceptual developments, emphasizing how metagenomics has broadened detection capabilities and enhanced understanding of ARG-host associations and the ecological risks associated with resistance genes. While previous reviews have addressed individual aspects of ARG research in environmental contexts, such as sequencing strategies or database development, few have integrated these components within a unified framework. This review addresses this gap by integrating ARG detection, host identification methods, quantification strategies, and risk assessment into a comprehensive synthesis. By consolidating these topics, the review offers a thorough overview of methodological and analytical advances in ARG investigation within environmental research.

Metagenomic sequencing methods: from second-generation to third-generation technologies

The development of DNA sequencing technologies has significantly deepened the understanding of microbial taxonomy and functions in complex environmental samples. Since the introduction of first-generation sequencing by Frederick Sanger in 1977, the field has undergone a series of technological revolutions<sup>[30]</sup>. For example, the emergence of second- and third-generation sequencing over the past decades has significantly reduced sequencing costs while markedly improving throughput and speed (Table 1). These advances have made large-scale genomic and metagenomic studies feasible on an unprecedented scale. Nowadays, both second- and third-generation sequencing technologies are widely used to study antibiotic-resistant bacteria (ARB) and ARGs<sup>[17,31]</sup>.

Second-generation sequencing technologies

Second-generation sequencing technologies marked a critical paradigm shift from the linear, low-throughput nature of traditional methods to massively parallel high-throughput approaches<sup>[32]</sup>. Second-generation sequencing introduced massively parallel DNA

Table 1 Comparison of commonly used sequencing platforms

Sequencing Technology	Platform	Read length	Bias	PCR-free	Real-time sequencing	Base modification detection	Accuracy	Cost
Second generation	Illumina DNBSEQ	~150 to 300 bp	PCR-related bias	No	No	No	> 99.9%	Low
		~150 to 300 bp	PCR-related bias	No	No	No	> 99.9%	Low
Third generation	PacBio SMRT	Tens of kb	Low bias due to single-molecule sequencing	Yes	No	Yes	> 99.9%	High
	ONT Nanopore	Tens of kb	Signal fluctuations	Yes	Yes	Yes	> 99.75%	High

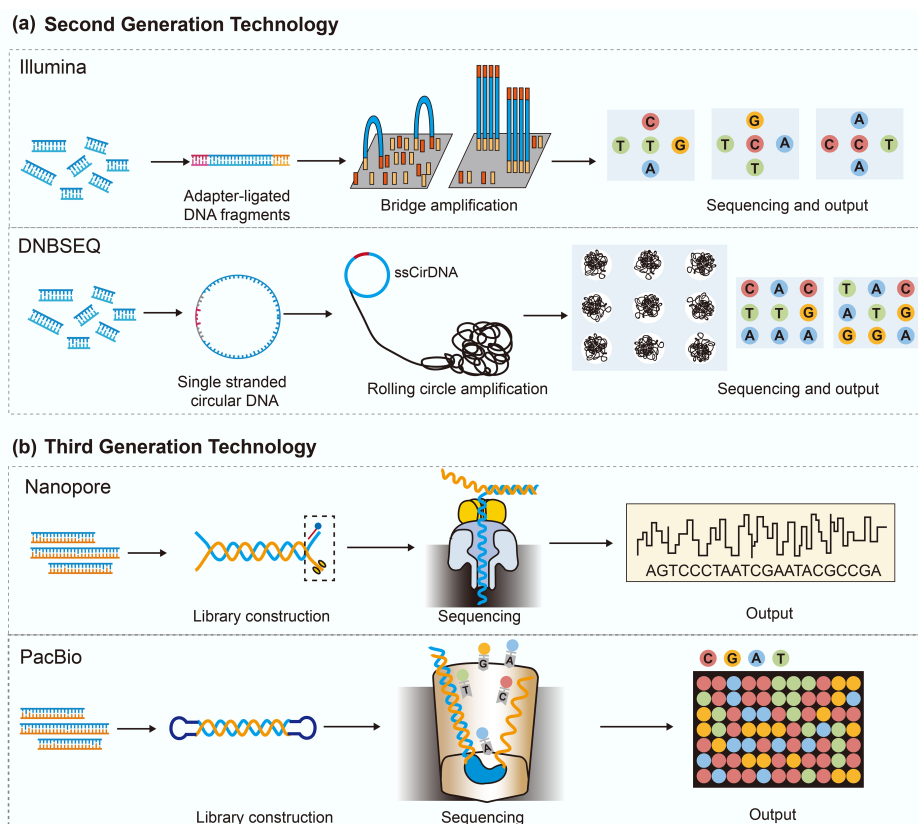
sequencing, allowing millions of fragments to be read simultaneously and significantly increasing sequencing throughput<sup>[33]</sup>. In contrast to labor-intensive, time-consuming Sanger sequencing, second-generation sequencing technology can complete whole-genome sequencing within days or even hours. The workflow of second-generation sequencing usually involves several steps, including fragmentation, tagging, and amplification<sup>[34]</sup>. Specifically, the target DNA is first broken into short fragments. These fragments are then tagged with adapters, followed by bridge PCR or cluster amplification to create sequencing templates (Fig. 1a). The actual sequencing is performed using reversible terminators or other chemical methods to identify the bases.

Second-generation sequencing is currently the most widely used sequencing technology, sequencing platforms including 454 pyrosequencing (Roche, Basel, Switzerland), SOLiD sequencing (Applied Biosystems, Foster City, CA, USA), Illumina sequencing-by-synthesis (Illumina, San Diego, CA, USA), Ion Torrent semiconductor sequencing (Thermo Fisher Scientific, Waltham, MA, USA), and DNA nanoball sequencing (DNBSEQ, Beijing Genomics Institute/MGI Tech Co., Shenzhen, China). Among these, Illumina's sequencing-by-synthesis and BGI's DNBSEQ platforms have become the dominant technologies owing to their high throughput, accuracy, and cost efficiency. Several distinct advantages support its widespread use<sup>[35]</sup>. First, it provides high sequencing throughput, capable of processing millions of DNA fragments in a single run, thereby enabling large sample batches to be sequenced within days or even hours. Second, the per-base cost of sequencing has been significantly reduced by continuous technological advancements and increasing commercial competition, making comprehensive sequencing accessible to a wide range of laboratories. Third, second-generation sequencers deliver high accuracy for short reads and can detect low-frequency

variants with high accuracy and depth. Fourth, second-generation sequencing workflows are highly automated, with most steps performed on automated instruments or liquid-handling platforms. However, second-generation technologies also have notable limitations. A primary drawback of second-generation sequencing technology is its short read length (typically 150–300 base pairs), which makes it difficult to resolve complex genomic regions. In addition, PCR-based clonal amplification of DNA libraries can introduce biases and errors, with some fragments amplifying more efficiently than others, leading to uneven coverage. Lastly, second-generation sequencing results in massive datasets, posing substantial demands on data processing, storage, and interpretation<sup>[32,36,37]</sup>.

### Third-generation sequencing technologies

The application of third-generation sequencing has expanded, serving both as a supplement to and a replacement for second-generation methods<sup>[34]</sup>. At present, third-generation sequencing is dominated by two platforms, Single-Molecule Real-Time (SMRT) sequencing from Pacific Biosciences and nanopore sequencing from Oxford Nanopore Technologies (ONT)<sup>[38]</sup> (Fig. 1b). SMRT sequencing relies on the real-time observation of nucleotide incorporation by a DNA polymerase immobilized at the base of a zero-mode waveguide<sup>[39]</sup>. As deoxynucleotide triphosphates (dNTPs) are added to the growing DNA strand, their fluorescent labels emit base-specific signals that are detected with high temporal resolution. The identity of each incorporated nucleotide is inferred based on the distinct spectral and kinetic properties of these fluorescent emissions<sup>[40]</sup>. In contrast, nanopore sequencing detects DNA molecules as they pass through membrane-embedded nanopores under an electric field, allowing direct, real-time reading of nucleotide sequences. The passage of nucleotides through



**Fig. 1** The principle of second- and third-generation sequencing technologies.

the nanopore produces characteristic fluctuations in the ionic current, which can be decoded to identify the DNA sequence<sup>[41]</sup>. These changes are sequence-dependent and are decoded in real time to reconstruct the underlying nucleotide sequence. Furthermore, several companies have developed platforms based on third-generation sequencing technologies, including Axio Biotechnology, Qi-Tan Gene Sequencing Pioneer, Beijing Polyseq Biotech, and BGI.

Third-generation sequencing offers notable improvements over second-generation sequencing<sup>[42,43]</sup>. First, third-generation sequencing can produce long reads spanning several kilobases, significantly improving the resolution of complex genomic regions. Second, third-generation sequencing workflows are PCR-free, enabling the direct sequencing of native DNA molecules and thereby reducing amplification-associated biases and errors. Third, nanopore sequencing provides real-time access to sequencing information during the run, facilitating rapid turnaround, adaptive sampling, and time-sensitive decisions. Lastly, third-generation sequencing can detect base modifications, such as DNA methylation, allowing simultaneous capture of both genetic and epigenetic information from the same molecule. However, third-generation sequencing technologies are still limited by several technical and financial factors. First, the sequencing cost per base of third-generation sequencing remains higher than that of second-generation sequencing. Second, third-generation sequencing platforms generally exhibit higher raw error rates, thereby requiring consensus sequencing or hybrid correction strategies. Third, their data throughput per run is comparatively lower, which can be a constraint in applications requiring deep coverage, such as metagenomic profiling. Finally, long read data present unique analytical challenges, including error correction and structural variant detection, which require specialized bioinformatics tools and workflows.

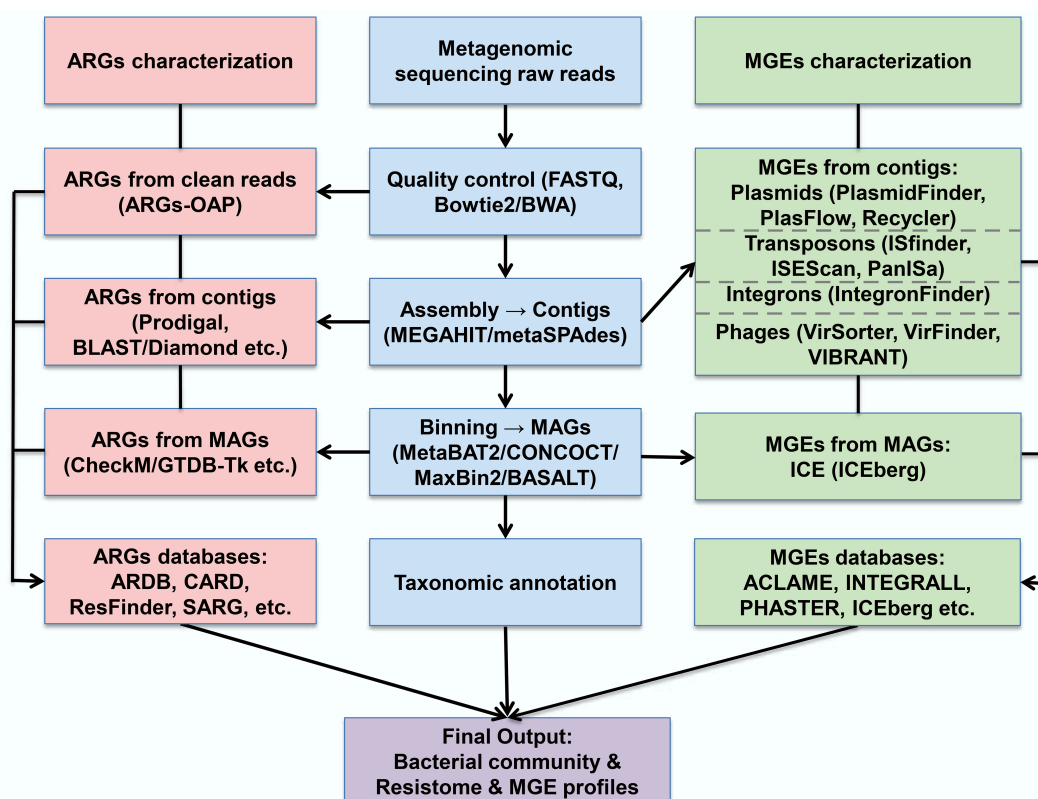
## Characterization of ARGs and MGEs based on a metagenomics approach

In recent years, metagenomics has dramatically improved the identification of ARGs and MGEs, providing crucial insights into the occurrence and dissemination of antibiotic resistance<sup>[44,45]</sup> (Fig. 2). The second-generation sequencing technology has become the mainstream platform for metagenomic research due to its high throughput, high accuracy, and relatively low cost<sup>[46]</sup>. Furthermore, third-generation sequencing technology, i.e., long-read sequencing, has significantly expanded the scope of metagenomic analysis<sup>[47,48]</sup>.

### Metagenomics data analysis workflow

In studies using the metagenomics method for the detection of ARGs and MGEs, bioinformatics is critical. The initial step involves quality control of raw sequencing reads, including the removal of low-quality reads, trimming of adapter sequences, and filtering for potential contaminants<sup>[49]</sup>. In bacterial metagenomic analysis, host-derived reads are also commonly removed by aligning them to a reference host genome using tools such as Bowtie2 or BWA<sup>[50,51]</sup>. Once quality control and host removal are completed, the clean reads are assembled *de novo* into contiguous sequences (contigs). Common assemblers used in metagenomic contexts include MEGAHIT and metaSPAdes, both of which are optimized for the complexity and uneven coverage of environmental microbial communities<sup>[24,52]</sup> (Fig. 3).

Assembly enables the reconstruction of longer genomic fragments, facilitating the detection of novel genes and gene clusters that are difficult to identify from unassembled short reads. This step significantly enhances the ability to characterize the diversity and genomic context of ARGs and MGEs in complex microbial environments. Subsequently, contigs can be grouped into metagenome-



**Fig. 2** Workflow for metagenomic-based identification of ARGs and MGEs.



assembled genomes (MAGs) using binning algorithms such as BASALT, MetaBAT2, CONCOCT, or MaxBin2<sup>[23,53–55]</sup>. These MAGs provide draft-level genomes that support downstream functional annotation, taxonomic classification, and comparative genomic analyses, offering a more complete picture of the resistome and mobilome architecture.

## Metagenomic characterization of ARGs and ARG abundance units

### ARG profiling from metagenomic data

In metagenomic studies of microbial communities, detecting ARGs relies on sequence comparisons against curated reference databases. Depending on the research objectives, alignments can be performed using short reads, assembled contigs, or metagenome assembled genomes. For analysis based on contigs and MAGs, open reading frames (ORFs) are first predicted from assembled sequences, and candidate ARGs are identified using sequence similarity and alignment criteria. Matches that meet the required identity and coverage criteria are annotated as ARGs utilizing a variety of established databases and tools. These databases differ in gene coverage, update frequency, classification strategy, and underlying curation methods.

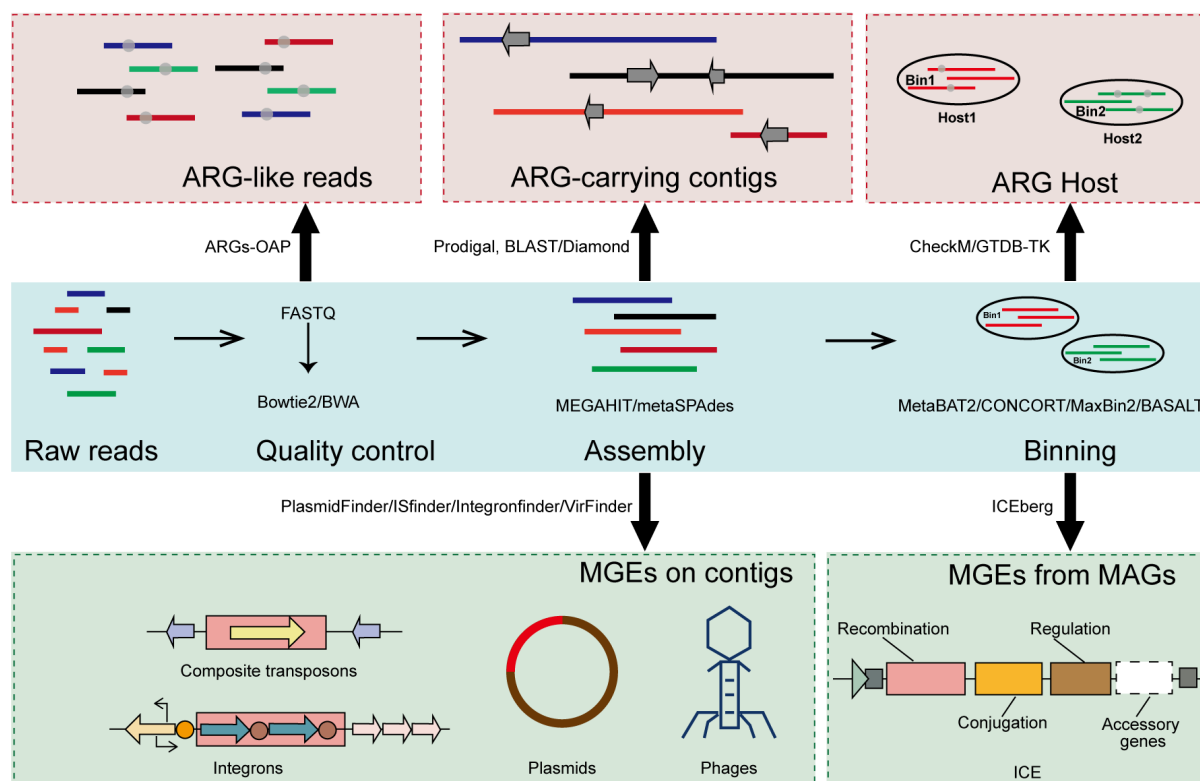
ResFinder is one of the earliest ARG databases, maintained by the Center for Genomic Epidemiology and focusing on acquired resistance genes in clinically relevant bacteria<sup>[56]</sup>. It employs BLAST for homology searches and performs well on assembled draft genomes from cultured isolates<sup>[57]</sup>. However, its gene coverage is relatively limited, which can be a drawback for complex or environmental samples. The Comprehensive Antibiotic Resistance Database (CARD) provides a more comprehensive, functionally annotated set of ARGs<sup>[58]</sup>. It includes annotations on resistance mechanisms, drug classes, and functional characteristics. When paired with the Resistance Gene Identifier (RGI) tool, CARD supports both

homology-based and model-based detection, providing a reliable framework for studies of clinical and environmental resistomes<sup>[59]</sup>. Its frequent updates and structured ontology enhance its reliability for high-confidence identification. For studies focusing on environmental samples, the Structured Antibiotic Resistance Gene (SARG) database provides a valuable framework<sup>[60]</sup>. It incorporates entries from CARD and the Antibiotic Resistance Genes Database (ARDB) and reclassifies ARGs into a unified three-level hierarchy (type, subtype, and reference sequence), facilitating standardized comparative analyses across datasets. The accompanying ARGs-OAP pipeline integrates gene prediction, quality filtering, and abundance normalization, and is widely used for resistome profiling in environmental metagenomics. Later, the CARD database also incorporated ARDB content, expanding its coverage<sup>[61]</sup>.

In addition to the databases described above, several other ARG databases and tools have been applied in various contexts, including ARDB, DeepARG, and AMRFinder<sup>[26,62,63]</sup>. While these databases are not discussed in detail here, they serve as valuable alternatives.

### ARG abundance units in the metagenomic method

Metagenomic sequencing is widely used to quantify ARGs in samples from various environments, and ARG abundance has been reported in multiple units across studies. Parts per million (PPM), originally introduced by Yang et al., was among the earliest units used to quantify ARG abundance and represents the number of ARG-like sequences per one million metagenomic reads<sup>[64]</sup>. Although PPM offers a straightforward way to describe relative abundance, but it does not account for gene length. Therefore, some studies have employed normalized units such as Reads Per Kilobase per Million reads (RPKM), which adjust read counts based on both gene length and sequencing depth<sup>[65]</sup>. This method allows comparisons between ARGs of different sizes, offering a more refined representation of relative abundance. However, both PPM and RPKM remain relative measures and do not



**Fig. 3** Overview of the metagenomic pipeline for identifying ARGs and MGEs.

directly reflect the number of ARG copies per microorganism or per cell. Therefore, more biologically meaningful units have been proposed<sup>[66]</sup>. A commonly adopted approach is to express ARG abundance as the number of copies per 16S rRNA gene, providing an estimate of ARG prevalence per bacterium and allowing direct comparison with those obtained via qPCR in other studies<sup>[19]</sup>. A further refinement is to normalize ARG counts to the number of microbial cells, yielding a unit of copies per cell, which allows for more direct ecological and risk-related interpretations<sup>[60,66]</sup>.

### **Absolute quantification of ARGs via the metagenomic method**

As mentioned above, metagenomic sequencing provides relative abundance data for ARGs, rather than direct measurement of absolute abundance. While relative abundance is useful for within-sample comparisons, it is often unreliable for comparing across samples. Therefore, absolute quantification has become increasingly important for accurately assessing the burden and dynamics of ARGs across samples. To address this, additional experimental measurements are combined with metagenomics to quantify absolute abundance. Quantitative PCR (qPCR) is among the most commonly used techniques. It can estimate the total number of 16S rRNA gene copies per sample<sup>[67]</sup>. These values are then used to normalize the sequencing-based ARG counts, yielding estimates of gene copy numbers per unit volume. However, these indirect approaches rely on multiple experimental platforms and are susceptible to variability introduced by differences in amplification efficiency or sample processing protocols.

To overcome these limitations, the spike-in internal standard method has been developed for the direct quantification of ARGs<sup>[68–70]</sup>. This approach involves adding a known quantity of exogenous materials to each sample. There are two types of internal standard, namely synthetic DNA fragments and cellular internal standards (for example, Gram-positive and Gram-negative bacteria). DNA fragment standards are typically introduced after DNA extraction but before library preparation, to correct for variations in library construction and sequencing depth<sup>[69]</sup>. In contrast, cellular internal standards are added before DNA extraction, allowing correction for biases in cell lysis and DNA recovery efficiency<sup>[70]</sup>. The internal standard sequence must not be present in the original sample, as confirmed by PCR. After spiking, a defined number of internal standard copies are added, and the sample is processed and sequenced alongside the native microbial DNA. During downstream analysis, sequencing reads are mapped to both the internal standard reference and the ARG database. By comparing the read counts of the internal standard with its known input amount, a conversion factor can be calculated, allowing inference of the absolute copy number of ARGs in the sample. These methods improve quantification accuracy by controlling for variation in library construction and sequencing depth, and enables more reliable comparisons of ARG loads across different samples and studies.

### **Characterization of mobile genetic elements (MGEs)**

Horizontal gene transfer (HGT) is the primary pathway for ARG dissemination and is mediated by mobile genetic elements, including plasmids, transposons, phages, integrons, and ICEs. Identifying MGEs in metagenomic datasets is essential for elucidating the mechanisms underlying ARG dissemination. There are different strategies and tools for identifying various MGEs (Fig. 3).

Plasmids are extrachromosomal DNA molecules that often carry multiple resistance genes and can move between different bacterial species<sup>[71]</sup>. The identification of plasmids from metagenomic data relies on a combination of computational tools and curated databases. Alignment-based methods, such as PlasmidFinder, detect

known plasmid replicon sequences through BLAST<sup>[72]</sup>. These methods offer high precision when matching known replicon types but are limited by database coverage and the fragmented nature of contigs in complex samples. In contrast, sequence composition-based approaches such as PlasFlow and cBar use machine learning models to classify DNA sequences as plasmid- or chromosome-derived<sup>[73,74]</sup>. Assembly-based methods like PlasmidSPAdes reconstruct putative plasmid sequences by identifying subgraphs with uneven coverage or circularity features, while Recycler specifically targets circular contigs indicative of closed plasmids<sup>[75,76]</sup>.

Composite transposons, which often mediate the horizontal transfer of ARGs, are mobile genetic elements composed of one or more cargo genes flanked by insertion sequences, i.e., IS elements. Most detection approaches identify transposase genes and flanking repeats by either directly aligning sequences to databases such as ISfinder or by applying profile-based tools such as ISEScan, which employs HMMs built from curated IS protein families<sup>[77,78]</sup>. Alternatively, read-mapping tools like panISa look for signs of novel insertions by identifying split or anomalous read pairs<sup>[79]</sup>. However, the fragmented nature of short-read assemblies often prevents the reconstruction of large complete composite elements.

Integrons are genetic elements that are capable of capturing and expressing gene cassettes, often carrying ARGs. They are identified by the presence of the integrase gene (*intI*) and the associated recombination site (*attC*). A widely used tool for this purpose is IntegronFinder<sup>[80]</sup>. It combines HMM searches for integrase proteins with covariance models that detect *attC* structures, enabling the identification of full-length integrons and incomplete forms, such as integrase-only or cassette-only arrays. Its design allows the application to isolate both genomes and complex metagenomes.

Bacteriophages contribute to the spread of ARGs by mediating transduction. However, their identification in metagenomic datasets remains challenging due to the lack of conserved marker genes and high sequence variability. To overcome this, homology-based tools such as VirSorter and VirSorter2 have been developed to scan metagenomic contigs for known viral hallmark genes<sup>[81]</sup>. These tools are effective for identifying well-characterized viral sequences and generally yield high precision, though they may miss novel or highly divergent phages. In contrast, k-mer-based methods such as VirFinder, DeepVirFinder, and Seeker rely on oligonucleotide frequencies or on deep learning models trained on viral genomic signatures<sup>[82,83]</sup>. These methods tend to have higher sensitivity for novel phages, but are also more prone to false positives, especially when host-derived sequences share similar sequence features. More recently, geNomad has emerged as a widely used tool for viral identification. Built on a foundation model framework, geNomad integrates gene content information and deep neural network representations, using deep-learning-derived protein embeddings for sequence classification and HMM-based profile searches for functional gene annotation<sup>[84]</sup>. Furthermore, CheckV has become the standard tool for downstream quality and completeness assessment, ensuring reliable interpretation of viral genomes<sup>[85]</sup>. Hybrid tools like VIBRANT and viralVerify incorporate both gene content and structural genome features, such as coding density and genome length, to improve classification accuracy<sup>[86]</sup>.

### **Approaches for identifying hosts of ARGs**

Metagenomic techniques have significantly expanded the understanding of ARGs and MGEs in both environmental and clinical settings. However, linking these elements to their microbial hosts remains a significant challenge. Determining the host microorganisms

of ARGs and MGEs is essential for elucidating their ecological distribution, assessing their potential for HGT, and evaluating public health risks. To address this gap, a variety of complementary approaches have been developed (Table 2), ranging from traditional cultivation and genome binning to single-cell and proximity-ligation techniques (Fig. 4)<sup>[87]</sup>.

## Cultured-based isolation method

The culture-based method is one of the earliest and most fundamental approaches for identifying antibiotic-resistant bacteria. This method recovers resistant bacteria by cultivating environmental and clinical samples on selective media containing specific antibiotics<sup>[88]</sup> (Fig. 4a). Isolates obtained through this process can subsequently be characterized using whole-genome sequencing or other molecular techniques to determine species identity as well as the presence of ARGs and MGEs<sup>[88]</sup>. The culture-based method could offer high resolution and specificity, with the added advantage of linking genotypic data to phenotypic resistance traits. Moreover, this allows for the direct association of ARGs with their native genomic and mobile contexts. However, most environmental bacteria remain unculturable under standard laboratory conditions. As a result, these methods often capture only a small and potentially biased subset of the resistome. Additionally, culturing is time-consuming and labor-intensive, and the selective conditions may favor the growth of strains that are not representative of the broader microbial community.

## Network analysis for host-gene association inference

Network analysis infers potential ARG hosts by calculating statistical co-occurrence patterns across multiple samples<sup>[19,89]</sup> (Fig. 4b). In this approach, the relative abundances of ARGs and microbial taxa are quantified, and pairwise correlations are calculated to identify associations between them. A co-occurrence network is then constructed, with nodes representing ARGs or taxa and edges indicating statistically significant correlations, suggesting possible host-gene relationships. While this method can identify candidate associations at scale, it is inherently indirect and does not provide definitive evidence of physical linkage. Additionally, complex microbial communities and the presence of low-abundance taxa may introduce noise and spurious correlations, limiting the reliability of inferred associations. Therefore,

network analysis is best used in combination with other methods that provide direct host-gene linkage.

## Emulsion paired isolation and concatenation PCR (epicPCR)

EpicPCR is a culture-independent method for directly associating functional genes with phylogenetically marked genes from the same microbial cell<sup>[90]</sup>. In this approach, individual cells are encapsulated within polyacrylamide beads and lysed inside microdroplets. A subsequent fusion PCR step amplifies both a portion of the 16S rRNA gene and a target ARG, generating a chimeric amplicon that links the functional gene to the taxonomic identity of its host. Sequencing these chimeric products provides high-resolution host-ARG associations without the need for genome assembly (Fig. 4c). EpicPCR is a high-throughput method that can process millions of droplets in parallel<sup>[91]</sup>. However, as a PCR-based method, it is subject to limitations such as primer mismatch, nonspecific amplification, and amplification bias across taxa. Additionally, more than one cell may be encapsulated within a single droplet in some single-cell or droplet-based assays, which can lead to false positives. The method is semi-quantitative, as variations in cell lysis efficiency and PCR performance can affect the relative representation of different taxa. Moreover, the detection range is constrained by primer design, limiting the approach to predefined target genes. Despite these limitations, epicPCR offers a powerful tool for mapping gene–host relationships in complex microbial communities<sup>[91,92]</sup>.

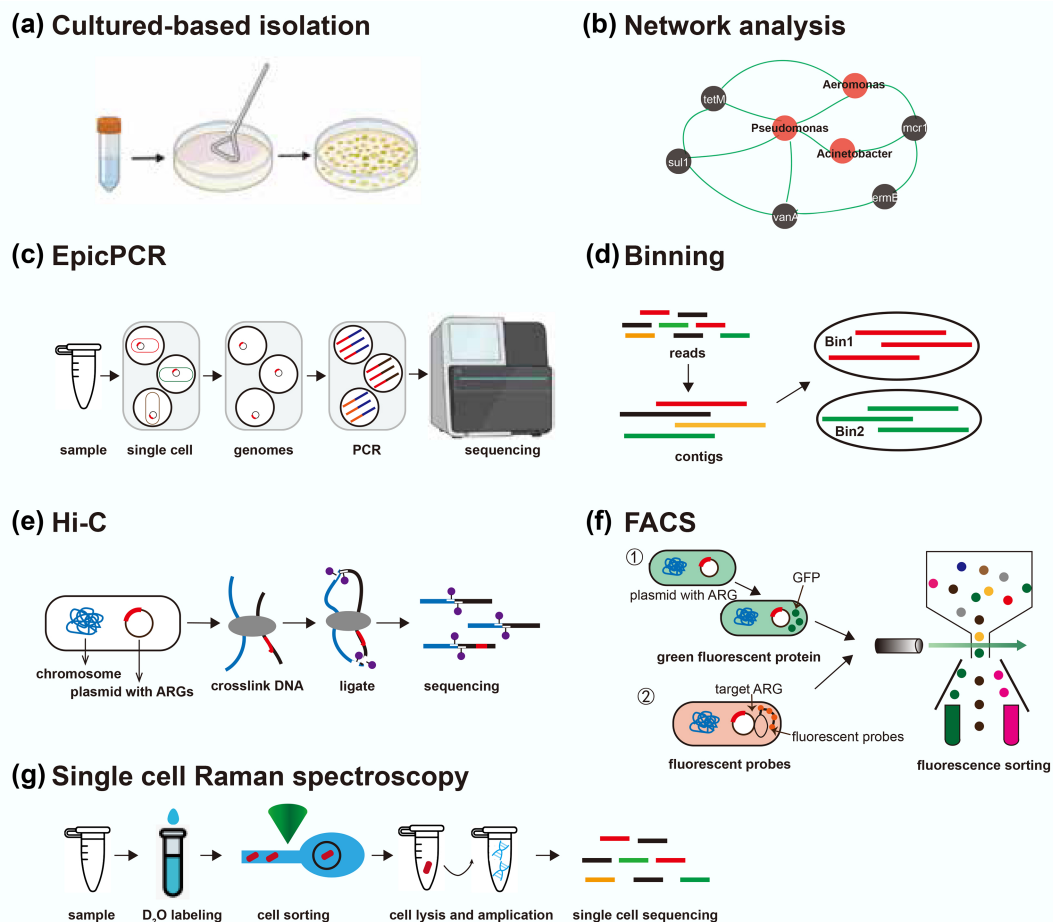
## Genome binning approach

For the binning-based approach, short reads are first assembled into contigs, which are then clustered into MAGs using sequence composition and abundance profiles across multiple samples<sup>[45]</sup> (Fig. 4d). By assigning a taxonomy to each MAG, any ARG found on a MAG can be linked to that host<sup>[93]</sup>. This approach allows recovery of a near-complete genome, including both chromosomal and plasmid sequences, thus preserving the genomic context of ARGs. MAGs also provide high-resolution taxonomic assignments and can reveal the co-occurrence of multiple ARGs within the same genome, offering valuable insights into multidrug resistance and gene linkage<sup>[94]</sup>. However, the binning-based method is limited by its reliance on assembly, which typically utilizes only a fraction of the total

**Table 2** Comparison of primary methods for identifying the hosts of ARGs and MGEs

Method	Principle	Advantages	Limitations
Culture-based isolation	Selective cultivation of resistant bacteria on antibiotic-containing media	High specificity; linking genotype to phenotype; preservation of native genomic and mobile context	Most bacteria unculturable; labor-intensive, and biased toward fast-growing strains
Network analysis	Statistical correlation of ARG abundance with microbial taxa abundance across samples	Culture-independent; predicting large-scale candidate associations	Indirect; spurious correlations; low-abundance taxa introduce noise; no physical linkage
EpicPCR	Droplet encapsulation of single cells; fusion PCR linking the ARG fragment with the 16S rRNA gene	Culture-independent; direct gene–host linkage	PCR bias; primer limitations; multi-cell droplets cause false positives;
Binning	Short reads are first assembled into contigs, which are then clustered into MAGs	Culture-independent; providing genomic context; linking ARGs to host genomes	ARG regions fragmentation; assembly bias; mis-binning; inability to link plasmids to host genomes
Hi-C	Cross-linking DNA in cells; proximity ligation yields read pairs linking contigs from the same cell	Culture-independent; physical linkage capture; linking ARG-carrying plasmid to host genomes; no prior targeting	Complex pretreatment; specialized bioinformatics; variable cross-linking efficiency; high sequencing-depth requirements
FACS	Flow cytometric sorting of fluorescently labeled bacteria	Culture-independent; isolation of target cells with high purity; support for conjugation studies	Requires fluorescent labeling; low throughput
Single cell Raman spectroscopy	Isotopic probing (D <sub>2</sub> O) identifies metabolically active ARB via Raman spectral shifts	Culture-independent; label-free operation; identify active ARB	Low throughput; low accessibility of single-cell Raman platforms

EpicPCR refers to emulsion paired isolation and concatenation PCR; Hi-C refers to high-throughput chromosome conformation capture; FACS refers to fluorescence-activated cell sorting.



**Fig. 4** Overview of current approaches for identifying hosts of ARGs. EpicPCR refers to emulsion paired isolation and concatenation PCR; Hi-C refers to high-throughput chromosome conformation capture; FACS refers to fluorescence-activated cell sorting. This figure is modified from Rice et al.<sup>[87]</sup>.

sequencing reads. Conserved ARGs or repetitive elements often disrupt contig assembly, resulting in short or fragmented ARG-containing sequences that are difficult to bin accurately. Consequently, crucial contextual information may be lost, such as the taxonomic origin or mobility of a gene. Additionally, contigs can be misassigned to incorrect bins, producing chimeric MAGs, and mobile genetic elements such as plasmids may remain unbinned due to their distinct sequence characteristics<sup>[52]</sup>.

### High-throughput chromosome conformation capture (Hi-C)

Chromosome conformation capture (3C), and 3C-based methods, including Hi-C, were initially developed to study chromatin interactions in eukaryotic cells through proximity ligation. Hi-C begins with formaldehyde fixation of intact cells to cross-link DNA regions that are in close spatial proximity<sup>[95]</sup>. After DNA fragmentation and re-ligation, paired-end sequencing of the Hi-C library yields read pairs that link genomic regions originally co-localized within the same cell (Fig. 4e). In metagenomic applications, this linkage information can be combined with shotgun assembly to assign contigs, including plasmids or fragments containing ARGs, to specific microbial genomes<sup>[96]</sup>. Unlike traditional 3C approaches, which depend on predetermined primers to detect selected loci, Hi-C captures chromatin interactions without prior sequence targeting. However, the Hi-C technique requires complex sample preparation and specialized bioinformatics to interpret the interaction data, and cross-linking efficiency can vary, potentially

biasing the detection of DNA proximities. Additionally, resolving the genomes of low-abundance microorganisms in highly diverse communities from metagenomic Hi-C data can be challenging.

### Fluorescence-activated cell sorting (FACS)

Fluorescence-activated cell sorting (FACS) is a technique that integrates flow cytometry with downstream molecular analysis, such as metagenomics, to sort and isolate target cells based on fluorescence. In this approach, the target ARGs or ARB must be fluorescently labeled to be detected and isolated<sup>[97]</sup> (Fig. 4f). One advantage of FACS is the ability to physically separate and collect fluorescently tagged cells with high purity, facilitating targeted study of their genomic or phenotypic properties. For example, FACS has been applied to study the conjugative transfer of ARGs in environmental samples by sorting donor cells carrying labeled ARGs and recipient cells that acquire them during horizontal gene transfer<sup>[97]</sup>. Despite these strengths, FACS is not well-suited for comprehensive profiling of ARG hosts in complex microbial communities, because only pre-selected or labeled targets can be detected. Combining FACS with techniques such as rolling-circle amplification FISH (RCA-FISH) or catalyzed reporter deposition FISH (CARD-FISH) may enable *in situ* detection of ARG hosts and broaden their use in environmental microbiology. However, these combined approaches are constrained by low throughput, the high cost of fluorescent probes, and the need for specialized instrumentation and expertise.



## Single-cell Raman spectroscopy (SCRS)

Single-cell Raman spectroscopy is an emerging, culture-independent method for identifying ARB at single-cell resolution. Single-cell Raman spectroscopy is often combined with isotopic probing (e.g., D<sub>2</sub>O-labeled water) to distinguish metabolically active ARB cells from non-resistant or inhibited cells under antibiotic exposure<sup>[98]</sup> (Fig. 4g). Cells that grow and incorporate deuterium exhibit characteristic shifts in their Raman spectra (due to C-D bond formation), allowing resistant and metabolically active bacteria to be distinguished from susceptible or inactive cells under antibiotic exposure. SCRS combined with Raman-activated cell sorting (RACS) enables single-cell isolation based on Raman spectral profiles, enabling further analyses such as genome sequencing or ARG detection. A significant advantage of SCRS is that it does not require prior knowledge of specific resistance genes or the use of fluorescent labels, making it particularly valuable for detecting unknown or unculturable resistant taxa. However, the technique is limited by low throughput, complex instrumentation, and the inability to directly identify ARG sequences, necessitating its use in conjunction with complementary molecular methods for comprehensive characterization.

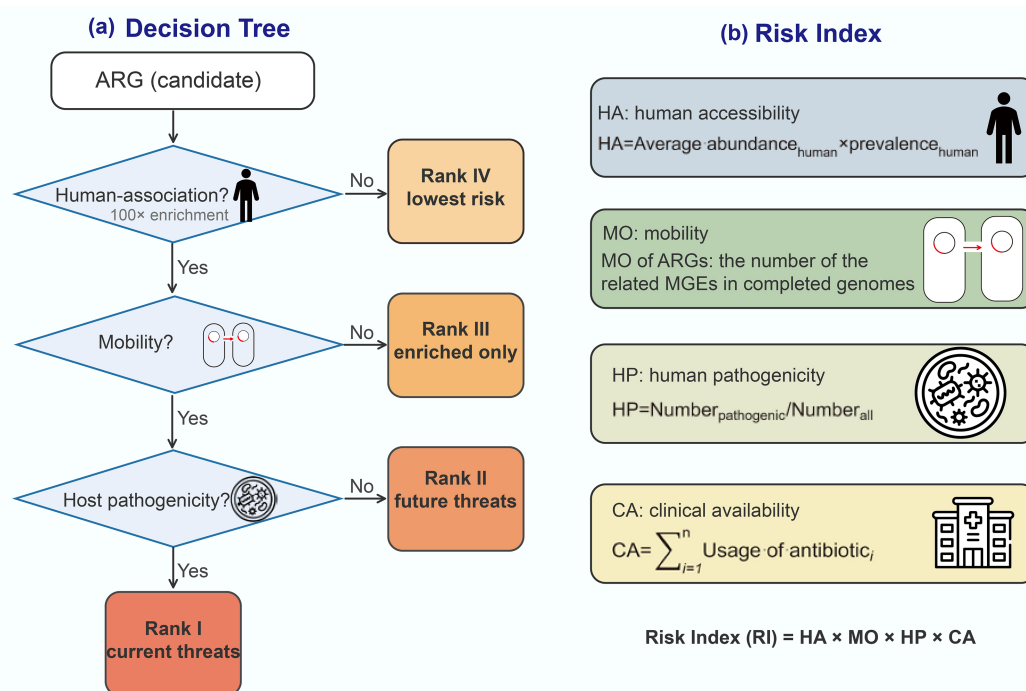
## Comprehensive risk assessment of ARGs

Metagenomic sequencing has significantly improved the ability to characterize ARGs profiles in diverse environments. However, not every resistance gene poses the same risk, and identifying which ARGs pose a threat to human health is not straightforward. Therefore, comprehensive risk assessment frameworks are needed to assess the threat posed by ARGs. In recent years, studies have reported ARG risk evaluation at two complementary levels: one focusing on the inherent risk of individual genes based on their genetic context, mobility, and host range; and the other assessing the overall resistome risk at the sample level by integrating abundance, potential for HGT, and pathogen association.

## Gene-level risk evaluation of ARGs

Recent methods score individual resistance genes using ecological and genomic criteria. Zhang et al. first reported a simple decision tree based on three key criteria: (1) human-association (enrichment in anthropogenic environments); (2) mobility (presence on mobile genetic elements); and (3) host pathogenicity (presence in known human pathogens)<sup>[99]</sup>. Using these criteria, ARGs are classified into four risk ranks<sup>[29]</sup> (Fig. 5a). ARGs that do not meet the first criterion are designated 'Rank IV' (lowest risk). Those meeting only enrichment are Rank III, those meeting enrichment and mobility (but not yet in pathogens) are Rank II ('future threats'), and those meeting all three are Rank I ('current threats'). This framework clearly differentiates high-risk ARGs (Rank I and II) from the lower-risk pool (Rank III and IV), and aligns well with expert assessments of clinical importance.

Other gene-centric approaches assign a continuous risk index rather than discrete ranks. For example, Zhang et al. define four quantitative indicators capturing different risk dimensions: human accessibility (HA), mobility (MO), human pathogenicity (HP), and clinical availability (CA)<sup>[99]</sup> (Fig. 5b). In this framework, HA is the average abundance multiplied by the prevalence of the ARG in human-associated samples; MO is the number of MGEs known to carry the ARG; HP is the fraction of ARG-hosting taxa that are human pathogens; and CA sums the clinical usage of antibiotics the ARG can resist. The gene-level risk index (RI) is then calculated as  $RI = HA \times MO \times HP \times CA$ , so that high values indicate that an ARG is abundant in humans, widely mobile, present in pathogens, and confers resistance to heavily used drugs. These gene-level frameworks are directly interpretable: each ARG is classified or scored for risk and can be monitored or reported as a priority. They rely on metagenomic surveillance data and reference databases to determine human enrichment, mobility, and presence in pathogen genomes. However, a limitation is that these assessments are based on sequence homology and co-occurrence data rather than on direct experimental confirmation of resistance phenotypes.



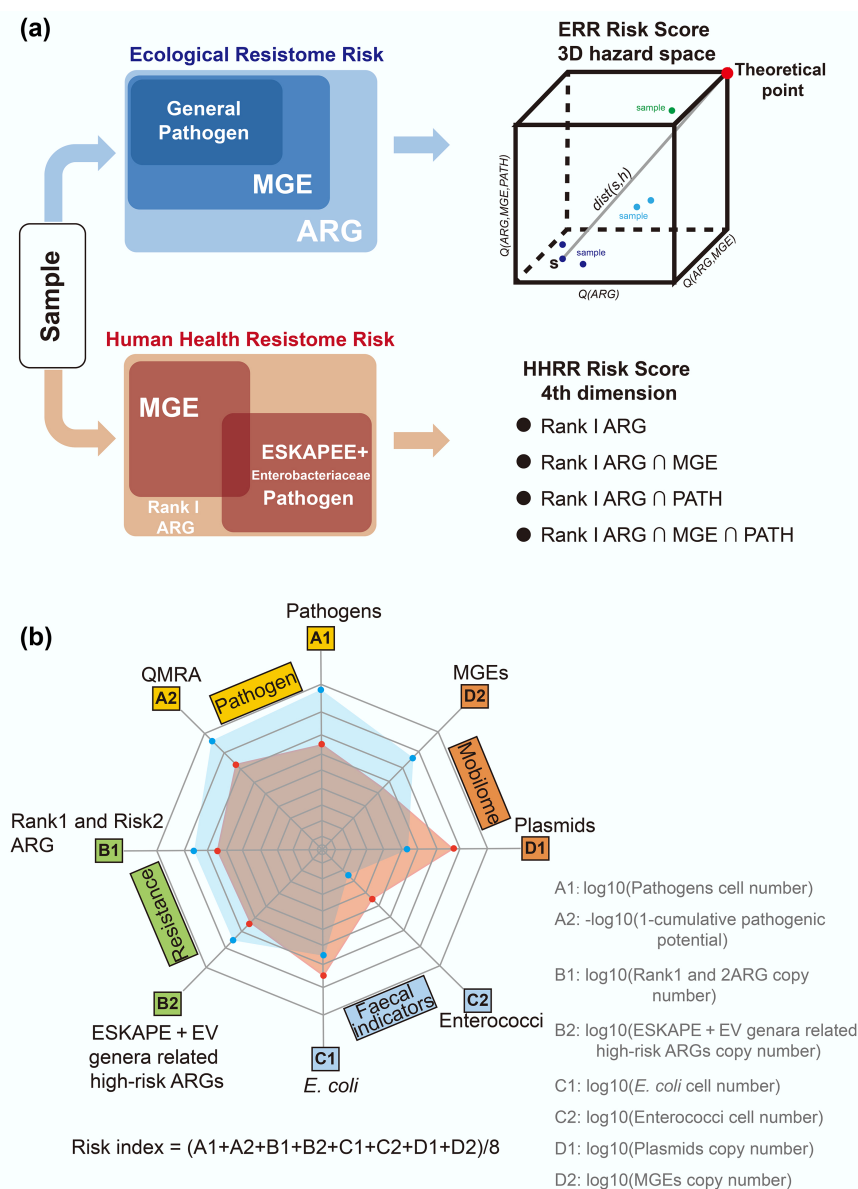
**Fig. 5** Gene risk assessment framework. (a) Assessment workflow for decision tree<sup>[29]</sup>. (b) Risk index calculation<sup>[99]</sup>.

## Sample-level risk evaluation of ARGs

Rather than evaluating individual genes, sample-level approaches assign a risk score to an entire metagenomic sample or environment, based on its collective resistome. These methods integrate gene occurrences and their context to rank environments or hosts by overall ARG threat. A well-known pipeline is MetaCompare, which estimates resistome risk by analyzing assembled metagenomic contigs for ARG presence and co-occurrence with mobility and pathogen markers<sup>[100]</sup> (Fig. 6a). The MetaCompare pipeline first *de novo* assembles the metagenomic sequencing reads, then screens the assembled contigs for three categories of features: (1) contigs containing any ARG; (2) contigs containing an ARG co-located with an MGE; and (3) contigs containing an ARG co-located with both an MGE and a pathogen-associated sequence. The proportions of contigs in each category are normalized, and each sample is placed in a three-dimensional hazard space, with samples assigned a resistome risk score. In effect, samples

with more ARGs co-localized with MGEs and pathogens rank as higher risk. The output is a single quantitative score for each sample, allowing comparison across environments or treatments.

Furthermore, Zhang et al. extended their gene-level index into a sample-level metric<sup>[99]</sup>. In their framework, once each ARG is assigned a gene RI as described above, the risk index of a sample is calculated as the abundance-weighted sum of the RIs of all ARGs in that sample (Fig. 6b). That means a sample's risk reflects both how many high RI genes it carries and how abundant they are. In principle, this approach yields a continuous risk index for each metagenome, leveraging the same underlying data (ARG annotations, abundances, MGE/host associations) computed in the gene-level analysis, but aggregating them to the sample level. Other sample-level metrics have been proposed. For instance, the MetaCompare pipeline has been updated to version 2.0 to distinguish between human health-related resistome risk and broader



**Fig. 6** Sample risk assessment framework. (a) Risk score assessment workflow based on MetaCompare 2.0<sup>[101]</sup>. (b) The risk assessment framework integrates four key components, including absolute quantification<sup>[70]</sup>. ARG refers to antibiotic resistance gene; MGE refers to mobile genetic element, PATH refers to pathogen; ESKAPE refers to *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter species*; EV refers to *Escherichia coli* and *Vibrio*.

ecological resistome risk for each sample<sup>[101]</sup>. In general, sample-level frameworks emphasize the mobility potential of ARGs, along with overall ARG load and the taxonomic composition of the community.

More recently, Shi et al. have developed a comprehensive framework for sample risk assessment that integrates four key components: pathogens (A), resistance (B), faecal indicators (C), and the mobilome (D)<sup>[70]</sup>. This framework quantifies multiple risk dimensions, including pathogens abundance (A1) and cumulative likelihood of infection (A2); absolute abundance of high-risk ARG categories such as rank 1, 2 (B1) and ESKAPE (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter species*) + EV (*Escherichia coli* and *Vibrio*)-related genes (B2); abundance of faecal indicators *E. coli* (C1) and *Enterococci* (C2); and the prevalence of plasmids (D1) and other MGEs (D2). An overall environmental risk index is then derived as the arithmetic mean of these eight criteria, with higher values indicating greater sample-level environmental risk.

It should be noted that the data requirements and outputs differ between gene- and sample-metric approaches. Gene-level schemes require large reference databases of ARGs, MGEs, and pathogen genomes, as well as broad metagenomic surveys to determine where each ARG is enriched. While read-based methods are suitable for gene-level profiling, assembly-based strategies are often employed to recover genomic context and improve annotation. Building on these, sample-level pipelines typically involve *de novo* assembly and integration with curated databases, making them computationally intensive. The interpretability of the results also differs: gene-level scores directly label specific ARGs as high-risk (facilitating targeted monitoring), whereas sample risk scores provide a single index per environment that must be contextualized (e.g., comparing hospitals vs soils). Both levels of assessment are complementary, and together they contribute to a more comprehensive understanding of antibiotic resistance risks in different environments.

## Quantification strategies and their implications for risk assessment

Several of the above-mentioned risk assessment frameworks rely on relative abundance derived from metagenomic sequencing. While relative metrics allow comparisons of ARGs within a sample, they have several limitations. First, changes in relative abundance may reflect shifts in other community members rather than an actual change in the absolute burden of resistance. Second, relative abundance values cannot be readily compared across studies or environments because microbial biomass and community composition vary among samples. As a result, risk scores derived solely from relative abundance may overestimate risks in low-biomass environments or underestimate risks in high-biomass settings where absolute ARG loads are high.

With the increasing application of absolute quantification approaches, some recent frameworks have begun to incorporate absolute ARG abundance into risk evaluation<sup>[70]</sup>. Accumulating evidence demonstrates that absolute quantification is essential for correcting compositional biases. Integrating absolute abundance into risk assessment captures not only the genetic attributes of ARGs but also their real-world concentrations, which more accurately reflect transmission potential and public health relevance. This distinction is critical because the ecological impact and public health risk posed by ARGs are ultimately determined by the absolute number of resistance genes, their hosts, and the biomass that carries them. For example, two samples may display similar relative ARG profiles yet

differ by orders of magnitude in their total microbial biomass; only absolute measurements can reveal whether an environment harbors a negligible number of resistance determinants or discharges large quantities of high-risk ARGs into downstream ecosystems.

## Summary and outlook

Metagenomic approaches have expanded the ability to detect, quantify, and characterize ARGs across diverse environments. They facilitate the identification of both known and emerging resistance genes, reveal their association with MGEs, and allow inferences about their microbial hosts. These insights are essential for understanding the mechanisms and pathways driving the dissemination of antibiotic resistance.

Looking ahead, several technological developments are likely to influence how ARGs are monitored and evaluated. Improvements in long-read and real-time sequencing are beginning to provide clearer resolution of plasmids, genomic islands, and other mobile elements that are often difficult to assemble from short reads. Advancements in host-linking technologies will allow more accurate assignment of ARGs to their native hosts. At the computational level, hybrid assembly pipelines, deep learning-based ARG prediction, and comprehensive risk modeling frameworks are expected to enhance the detection of novel resistance determinants and refine assessments of their mobility and pathogenicity. Additionally, progress in absolute quantification and standardized methodology will also help improve comparisons of ARGs across different studies and environments.

By integrating these technological advances with coordinated international monitoring and data sharing, future work can achieve more precise identification of high-risk ARGs, improve the design of targeted interventions, and ultimately support efforts to safeguard antibiotic effectiveness and public health.

## Author contributions

The authors confirm their contributions to the paper as follows: Peiju Fang: writing—original draft, validation, writing—review and editing; Zehui Yu: writing—original draft, validation, writing—review and editing; Jin Huang: validation, writing—review and editing; Bing Li: conceptualization, funding acquisition, supervision, validation, writing—review and editing. All authors reviewed the results and approved the final version of the manuscript.

## Data availability

The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

## Funding

This study was funded by the National Key R&D Program of China (Grant No. 2022YFE0103200), National Natural Science Foundation of China (Grant Nos 22176107 and 22476108), Shenzhen Science and Technology Innovation Bureau (Grant Nos SGDX20230821091559021 and GJHZ20240218113559006).

## Declarations

## Competing interests

The authors declare that they have no conflict of interest.

## Author details

<sup>1</sup>Institute of Environment and Ecology, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China; <sup>2</sup>Key Laboratory of Microorganism Application and Risk Control, Ministry of Ecology and Environment, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China; <sup>3</sup>School of Minerals Processing and Bioengineering, Central South University, Changsha 410083, China; <sup>4</sup>Key Laboratory of Urban Water Supply, Water Saving and Water Environment Governance in the Yangtze River Delta of Ministry of Water Resources, Shanghai 200092, China

## References

- [1] Macesic N, Uhlemann AC, Peleg AY. 2025. Multidrug-resistant Gram-negative bacterial infections. *The Lancet* 405:257–272
- [2] Naghavi M, Vollset SE, Ikuta KS, Swetschinski LR, Gray AP, et al. 2024. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *The Lancet* 404:1199–1226
- [3] Larsson DGJ, Flach CF. 2022. Antibiotic resistance in the environment. *Nature Reviews Microbiology* 20:257–269
- [4] Li L, Li B, Yin X, Xia Y, Yang Y, et al. 2025. Assessing antimicrobial resistance connectivity across One Health sectors. *Nature Water* 3:1100–1113
- [5] Hernando-Amado S, Coque TM, Baquero F, Martínez JL. 2019. Defining and combating antibiotic resistance from One Health and Global Health perspectives. *Nature Microbiology* 4:1432–1442
- [6] Soucy SM, Huang J, Gogarten JP. 2015. Horizontal gene transfer: building the web of life. *Nature Reviews Genetics* 16:472–482
- [7] Huang J, Yong H, Huang J, Che Y, Klümper U, et al. 2025. Microbial risks triggered by oral administration of antibiotics in fish aquaculture persist long after the legally mandated antibiotic withdrawal time. *Nature Water* 3:1057–1069
- [8] Martínez JL, Coque TM, Baquero F. 2015. What is a resistance gene? Ranking risk in resistomes. *Nature Reviews Microbiology* 13:116–123
- [9] Tang A, Zhang J, Huang J, Deng Y, Wang D, et al. 2024. Decrypting the viral community in aerobic activated sludge reactors treating antibiotic production wastewater. *Water Research* 265:122253
- [10] Zhang J, Tang A, Jin T, Sun D, Guo F, et al. 2024. A panoramic view of the virosphere in three wastewater treatment plants by integrating viral-like particle-concentrated and traditional non-concentrated metagenomic approaches. *iMeta* 3:e188
- [11] Cen T, Zhang X, Xie S, Li D. 2020. Preservatives accelerate the horizontal transfer of plasmid-mediated antimicrobial resistance genes via differential mechanisms. *Environment International* 138:105544
- [12] Jiao P, Zhou Y, Zhang X, Jian H, Zhang XX, et al. 2024. Mechanisms of horizontal gene transfer and viral contribution to the fate of intracellular and extracellular antibiotic resistance genes in anaerobic digestion supplemented with conductive materials under ammonia stress. *Water Research* 267:122549
- [13] Zhou H, Beltrán JF, Brito IL. 2021. Functions predict horizontal gene transfer and the emergence of antibiotic resistance. *Science Advances* 7:eabj5056
- [14] Yao Y, Maddamsetti R, Weiss A, Ha Y, Wang T, et al. 2022. Intra- and interpopulation transposition of mobile genetic elements driven by antibiotic selection. *Nature Ecology & Evolution* 6:555–564
- [15] Bengtsson-Palme J, Larsson DGJ, Kristiansson E. 2017. Using metagenomics to investigate human and environmental resistomes. *Journal of Antimicrobial Chemotherapy* 72:2690–2703
- [16] Li S, Zhang C, Li F, Hua T, Zhou Q, et al. 2021. Technologies towards antibiotic resistance genes (ARGs) removal from aquatic environment: a critical review. *Journal of Hazardous Materials* 411:125148
- [17] Nesme J, Cécillon S, Delmont TO, Monier JM, Vogel TM, et al. 2014. Large-scale metagenomic-based study of antibiotic resistance in the environment. *Current Biology* 24:1096–1100
- [18] Pillay S, Calderón-Franco D, Urhan A, Abeel T. 2022. Metagenomic-based surveillance systems for antibiotic resistance in non-clinical settings. *Frontiers in Microbiology* 13:1066995
- [19] Li B, Yang Y, Ma L, Ju F, Guo F, et al. 2015. Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. *The ISME Journal* 9:2490–2502
- [20] Beatson SA, Walker MJ. 2014. Tracking antibiotic resistance. *Science* 345:1454–1455
- [21] Rhoads A, Au KF. 2015. PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics* 13:278–289
- [22] Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, et al. 2022. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nature Methods* 19:823–826
- [23] Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, et al. 2014. Binning metagenomic contigs by coverage and composition. *Nature Methods* 11:1144–1146
- [24] Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676
- [25] McCorison CB, Kim T, Donato JJ, LaPara TM. 2025. Proximity-ligation metagenomic sequence analysis reveals that the antibiotic resistance makes significant transitions during municipal wastewater treatment. *Environmental Microbiology* 27:e70036
- [26] Arango-Argoty G, Garner E, Pruden A, Heath LS, Vikesland P, et al. 2018. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* 6:23
- [27] Yin X, Jiang XT, Chai B, Li L, Yang Y, et al. 2018. ARGs-OAP v2.0 with an expanded SARG database and Hidden Markov Models for enhancement characterization and quantification of antibiotic resistance genes in environmental metagenomes. *Bioinformatics* 34:2263–2270
- [28] dos Santos DFK, Istvan P, Quirino BF, Kruger RH. 2017. Functional metagenomics as a tool for identification of new antibiotic resistance genes from natural environments. *Microbial Ecology* 73:479–491
- [29] Zhang AN, Gaston JM, Dai CL, Zhao S, Poyet M, et al. 2021. An omics-based framework for assessing the health risk of antimicrobial resistance genes. *Nature Communications* 12:4765
- [30] Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, et al. 2017. DNA sequencing at 40: past, present and future. *Nature* 550:345–353
- [31] Boolchandani M, D'Souza AW, Dantas G. 2019. Sequencing-based methods and resources to study antimicrobial resistance. *Nature Reviews Genetics* 20:356–370
- [32] Hu T, Chitnis N, Monos D, Dinh A. 2021. Next-generation sequencing technologies: an overview. *Human Immunology* 82:801–811
- [33] Metzker ML. 2010. Sequencing technologies — the next generation. *Nature Reviews Genetics* 11:31–46
- [34] Buermans HPJ, den Dunnen JT. 2014. Next generation sequencing technology: advances and applications. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1842:1932–1941
- [35] Foox J, Tighe SW, Nicolet CM, Zook JM, Byrsk-Bishop M, et al. 2021. Performance assessment of DNA sequencing platforms in the ABRF Next-Generation Sequencing Study. *Nature Biotechnology* 39:1129–1140
- [36] Ari Ş, Arikian M. 2016. Next-generation sequencing: advantages, disadvantages, and future. In *Plant Omics: Trends and Applications*, eds Hakeem K, Tombuloğlu H, Tombuloğlu G. Cham: Springer. pp. 109–135 doi: [10.1007/978-3-319-31703-8\\_5](https://doi.org/10.1007/978-3-319-31703-8_5)
- [37] Mardis ER. 2008. Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics* 9:387–402
- [38] Cuber P, Chooneea D, Geeves C, Salatino S, Creedy TJ, et al. 2023. Comparing the accuracy and efficiency of third generation sequencing technologies, Oxford Nanopore Technologies, and Pacific Biosciences, for DNA barcode sequencing applications. *Ecological Genetics and Genomics* 28:100181
- [39] Larkin J, Henley RY, Jadhav V, Korlach J, Wanunu M. 2017. Length-independent DNA packing into nanopore zero-mode waveguides for low-input DNA sequencing. *Nature Nanotechnology* 12:1169–1175
- [40] Eid J, Fehr A, Gray J, Luong K, Lyle J, et al. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323:133–138



- [41] Petersen LM, Martin IW, Moschetti WE, Kershaw CM, Tsongalis GJ. 2019. Third-generation sequencing in the clinical laboratory: exploring the advantages and challenges of nanopore sequencing. *Journal of Clinical Microbiology* 58:e01315-19
- [42] Kumar KR, Cowley MJ, Davis RL. 2024. Next-generation sequencing and emerging technologies. *Seminars in Thrombosis and Hemostasis* 50:1026–1038
- [43] Liu L, Li Y, Li S, Hu N, He Y, et al. 2012. Comparison of next-generation sequencing systems. *BioMed Research International* 2012:251364
- [44] Chen B, Yuan K, Chen X, Yang Y, Zhang T, et al. 2016. Metagenomic analysis revealing antibiotic resistance genes (ARGs) and their genetic compartments in the Tibetan environment. *Environmental Science & Technology* 50:6670–6679
- [45] Djordjevic SP, Jarocki VM, Seemann T, Cummins ML, Watt AE, et al. 2024. Genomic surveillance for antimicrobial resistance — a One Health perspective. *Nature Reviews Genetics* 25:142–157
- [46] Ma L, Xia Y, Li B, Yang Y, Li LG, et al. 2016. Metagenomic assembly reveals hosts of antibiotic resistance genes and the shared resistome in pig, chicken, and human feces. *Environmental Science & Technology* 50:420–427
- [47] Che Y, Xia Y, Liu L, Li AD, Yang Y, et al. 2019. Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore metagenomic sequencing. *Microbiome* 7:44
- [48] Dai D, Brown C, Bürgmann H, Larsson DGJ, Nambi I, et al. 2022. Long-read metagenomic sequencing reveals shifts in associations of antibiotic resistance genes with mobile genetic elements from sewage to activated sludge. *Microbiome* 10:20
- [49] Nayfach S, Pollard KS. 2016. Toward accurate and quantitative comparative metagenomics. *Cell* 166:1103–1116
- [50] Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9:357–359
- [51] Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760
- [52] Uritskiy GV, DiRuggiero J, Taylor J. 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6:158
- [53] Qiu Z, Yuan L, Lian CA, Lin B, Chen J, et al. 2024. BASALT refines binning from metagenomic data and increases resolution of genome-resolved metagenomic analysis. *Nature Communications* 15:2179
- [54] Kang DD, Li F, Kirton E, Thomas A, Egan R, et al. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359
- [55] Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32:605–607
- [56] Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, et al. 2012. Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy* 67:2640–2644
- [57] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421
- [58] McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, et al. 2013. The comprehensive antibiotic resistance database. *Antimicrobial agents and chemotherapy* 57:3348–3357
- [59] Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, et al. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research* 48:D517–D525
- [60] Yang Y, Jiang X, Chai B, Ma L, Li B, et al. 2016. ARGs-OAP: online analysis pipeline for antibiotic resistance genes detection from metagenomic data using an integrated structured ARG-database. *Bioinformatics* 32:2346–2351
- [61] Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, et al. 2017. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Research* 45:D566–D573
- [62] Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, et al. 2019. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrobial Agents and Chemotherapy* 63:e00483-19
- [63] Liu B, Pop M. 2009. ARDB—antibiotic resistance genes database. *Nucleic Acids Research* 37:D443–D447
- [64] Yang Y, Li B, Ju F, Zhang T. 2013. Exploring variation of antibiotic resistance genes in activated sludge over a four-year period through a metagenomic approach. *Environmental Science & Technology* 47:10197–10205
- [65] Mao X, Yin X, Yang Y, Che Y, Xu X, et al. 2024. Standardization in global environmental antibiotic resistance genes (ARGs) surveillance. *Critical Reviews in Environmental Science and Technology* 54:1633–1650
- [66] Yin X, Chen X, Jiang XT, Yang Y, Li B, et al. 2023. Toward a universal unit for quantification of antibiotic resistance genes in environmental samples. *Environmental Science & Technology* 57:9713–9721
- [67] Li B, Li X, Yan T. 2021. A quantitative metagenomic sequencing approach for high-throughput gene quantification and demonstration with antibiotic resistance genes. *Applied and Environmental Microbiology* 87:e00871-21
- [68] Ju F, Beck K, Yin X, Maccagnan A, McArdell CS, et al. 2019. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. *The ISME Journal* 13:346–360
- [69] Satinsky BM, Gifford SM, Crump BC, Moran MA. 2013. Use of internal standards for quantitative metatranscriptome and metagenome analysis. In *Methods in Enzymology*. Volume 531. Amsterdam, the Netherlands: Elsevier. pp. 237–250 doi: 10.1016/B978-0-12-407863-5.00012-5
- [70] Shi X, Yang Y, Wang C, Xu X, Mao X, et al. 2025. Microbial risk assessment across multiple environments based on metagenomic absolute quantification with cellular internal standards. *Nature Water* 3:473–485
- [71] Wang X, Zhang H, Yu S, Li D, Gillings MR, et al. 2024. Inter-plasmid transfer of antibiotic resistance genes accelerates antibiotic resistance in bacterial pathogens. *The ISME Journal* 18:wrad032
- [72] Carattoli A, Zankari E, Garcia-Fernandez A, Larsen MV, Lund O, et al. 2014. *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrobial Agents and Chemotherapy* 58:3895–3903
- [73] Krawczyk PS, Lipinski L, Dziembowski A. 2018. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Research* 46:e35
- [74] Zhou F, Xu Y. 2010. cBar: a computer program to distinguish plasmid-derived from chromosome-derived sequence fragments in metagenomics data. *Bioinformatics* 26:2051–2052
- [75] Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, et al. 2016. plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics* 32:3380–3387
- [76] Rozov R, Brown Kav A, Bogumil D, Shterzer N, Halperin E, et al. 2017. Recycler: an algorithm for detecting plasmids from *de novo* assembly graphs. *Bioinformatics* 33:475–482
- [77] Siguier P, Pérochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Research* 34:D32–D36
- [78] Xie Z, Tang H. 2017. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* 33:3340–3347
- [79] Treepong P, Guyeux C, Meunier A, Couchoud C, Hocquet D, et al. 2018. *panLSa: ab initio* detection of insertion sequences in bacterial genomes from short read sequence data. *Bioinformatics* 34:3795–3800
- [80] Néron B, Littner E, Haudiquet M, Perrin A, Cury J, et al. 2022. Integron-Finder 2.0: identification and analysis of integrons across bacteria, with a focus on antibiotic resistance in Klebsiella. *Microorganisms* 10:700
- [81] Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, et al. 2021. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* 9:37

- [82] Ren J, Song K, Deng C, Ahlgren NA, Fuhrman JA, et al. 2020. Identifying viruses from metagenomic data using deep learning. *Quantitative Biology* 8:64–77
- [83] Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. 2017. VirFinder: a novel *k*-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* 5:69
- [84] Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, et al. 2024. Identification of mobile genetic elements with geNomad. *Nature Biotechnology* 42:1303–1312
- [85] Nayfach S, Camargo AP, Schulz F, Eloie-Fadrosch E, Roux S, et al. 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology* 39:578–585
- [86] Kieft K, Zhou Z, Anantharaman K. 2020. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* 8:90
- [87] Rice EW, Wang P, Smith AL, Stadler LB. 2020. Determining hosts of antibiotic resistance genes: a review of methodological advances. *Environmental Science & Technology Letters* 8:282–291
- [88] Garcia-Armisen T, Anzil A, Cornelis P, Chevreuil M, Servais P. 2013. Identification of antimicrobial resistant bacteria in rivers: insights into the cultivation bias. *Water Research* 47:4938–4947
- [89] Zhu YG, Zhao Y, Li B, Huang CL, Zhang SY, et al. 2017. Continental-scale pollution of estuaries with antibiotic resistance genes. *Nature Microbiology* 2:16270
- [90] Spencer SJ, Tamminen MV, Preheim SP, Guo MT, Briggs AW, et al. 2016. Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *The ISME Journal* 10:427–436
- [91] Liu S, Dai S, Deng Y, Li J, Zhang Y, et al. 2025. Long-read epicPCR enhances species-level host identification of clinically relevant antibiotic resistance genes in environmental microbial communities. *Environment International* 197:109337
- [92] Lou EG, Fu Y, Wang Q, Treangen TJ, Stadler LB. 2024. Sensitivity and consistency of long-and short-read metagenomics and epicPCR for the detection of antibiotic resistance genes and their bacterial hosts in wastewater. *Journal of Hazardous Materials* 469:133939
- [93] Zhao R, Yu K, Zhang J, Zhang G, Huang J, et al. 2020. Deciphering the mobility and bacterial hosts of antibiotic resistance genes under antibiotic selection pressure by metagenomic assembly and binning approaches. *Water Research* 186:116318
- [94] Liang J, Mao G, Yin X, Ma L, Liu L, et al. 2020. Identification and quantification of bacterial genomes carrying antibiotic resistance genes and virulence factor genes for aquatic microbiological risk assessment. *Water Research* 168:115160
- [95] Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, et al. 2012. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* 58:268–276
- [96] Stadler T, Press MO, Sullivan S, Liachko I, Top EM. 2019. Linking the resistome and plasmidome to the microbiome. *The ISME Journal* 13:2437–2446
- [97] Wang Y, Yu Z, Ding P, Lu J, Klümper U, et al. 2022. Non-antibiotic pharmaceuticals promote conjugative plasmid transfer at a community-wide level. *Microbiome* 10:124
- [98] Li HZ, Yang K, Liao H, Lassen SB, Su JQ, et al. 2022. Active antibiotic resistome in soils unraveled by single-cell isotope probing and targeted metagenomics. *Proceedings of the National Academy of Sciences of the United States of America* 119:e2201473119
- [99] Zhang Z, Zhang Q, Wang T, Xu N, Lu T, et al. 2022. Assessment of global health risk of antibiotic resistance genes. *Nature Communications* 13:1553
- [100] Oh M, Pruden A, Chen C, Heath LS, Xia K, et al. 2018. MetaCompare: a computational pipeline for prioritizing environmental resistome risk. *FEMS Microbiology Ecology* 94:fiy079
- [101] Rumi MA, Oh M, Davis BC, Brown CL, Juvekar A, et al. 2024. MetaCompare 2.0: differential ranking of ecological and human health resistome risks. *FEMS Microbiology Ecology* 100:fiac155



Copyright: © 2026 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.