2025-Enhanced detection for antibiotic resistance genes in wastewater samples using a CRISPR-enriched metagenomic method

2024 · 10.1101/2024.07.30.605462

8
Water Research 274 (2025) 123056
ELSEVIER
Contents lists available at ScienceDirect
Water Research
journal homepage: www.elsevier.com/ locate/watres
WATER
RESEARCH
A journal of the International New Association

Enhanced detection for antibiotic resistance genes in wastewater samples using a CRISPR-enriched metagenomic method

Yuqing Mao a,b,*, Joanna L. Shisler b,c, Thanh H. Nguyen a,b,d
a Department of Civil and Environmental Engineering, The Grainger College of Engineering, University of Illinois Urbana-Champaign, IL, USA
b Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, IL, USA
c Department of Microbiology, University of Illinois Urbana-Champaign, IL, USA
d Carle Illinois College of Medicine, University of Illinois Urbana-Champaign, IL, USA

ARTICLE INFO
Keywords:
Antibiotic resistance genes
Metagenomics
CRISPR
Wastewater

ABSTRACT
The spread of antibiotic resistance genes (ARGs) in the environment is a global public health concern. To date, over 5000 genes have been identified to express resistance to antibiotics. ARGs are usually low in abundance for wastewater samples, making them difficult to detect. Metagenomic sequencing and quantitative polymerase chain reaction (qPCR), two conventional ARG detection methods, have low sensitivity and low throughput limitations, respectively. We developed a CRISPR-Cas9-modified next-generation sequencing (NGS) method to enrich the targeted ARGs during library preparation. The false negative and false positive of this method were determined based on a mixture of bacterial isolates with known whole-genome sequences. Low values of both false negative (2/1208) and false positive (1/1208) proved the method's reliability. We compared the results obtained by this CRISPR-NGS and the conventional NGS method for six untreated wastewater samples. As compared to the ARGs detected in the same samples using the regular NGS method, the CRISPR-NGS method found up to 1189 more ARGs and up to 61 more ARG families in low abundances, including the clinically important KPC beta-lactamase genes in the six wastewater samples collected from different sources. Compared to the regular NGS method, the CRISPR-NGS method lowered the detection limit of ARGs from the magnitude of 10^{-4 to 10^{-5 as quantified by qPCR relative abundance. The CRISPR-NGS method is promising for ARG detection in wastewater. A similar workflow can also be applied to detect other targets that are in low abundance but of high diversity.

1. Introduction

Antibiotic resistance (AR) is an increasing threat to global public health. It is estimated that in the next decade, if without any intervention, AR will cause an annual loss of 3.4 trillion US dollars in gross domestic products and cause 24 million more people to live in extreme poverty (United Nations Environment

Programme, 2023). In addition to surveillance in clinical settings, AR surveillance based on wastewater has been conducted (Dong et al., 2024, Wang et al., 2024, Abia Akebe et al., 2022, Larsson and Flach, 2021, Lin et al., 2024, Wang et al., 2024, Mao et al., 2024, Auerbach et al., 2007). Culture-independent methods are commonly used for wastewater surveillance of AR (Liguori et al., 2022), because genes that encode AR are included in bacterial genomes. By analyzing the DNA extracted from wastewater samples, the AR profiles of the entire samples can be obtained. There are two challenges to detect antibiotic resistance genes (ARGs) in wastewater DNA samples. First, the concentration of many RGS can be very low because ARGs account for less than 1% of the DNA in bacterial genomes (Debroas and Siguret, 2019). Second, there is a wide range of ARG targets. To date, more than 5000 reference sequences of ARGs are recorded in The Comprehensive Antibiotic Resistance Database (CARD) (Alcock et al., 2023). Metagenomics and quantitative polymer chain reaction (qPCR) are two culture-independent methods that are widely used to detect ARGs (Liguori et al., 2022). These two methods have their pros and cons in ARG detection. qPCR has a high sensitivity, so it can capture the presence of ARGs that are in extremely low abundance in a sample (Bengsson-Palme et al., 2023). A previous study found that the detection limit of qPCR genes in manure and compost samples was 2 - 8 copies/mg, while that for metagenomics was around 3 \times 10^{4 copies/mg (Crossette et al., 2021). Another study also found that qPCR

could stably detect ARGs in water samples, while metagenomics could not constantly detect those in low abundances (Ferreira et al., 2023). Besides, qPCR can absolutely quantify the number of gene copies in a sample (Harshitha and Arunraj, 2021). However, a major drawback of qPCR is that it only returns fluorescent signals when short DNA sequence of 70 - 300 bp are detected (Ginzinger, 2002). These results can neither indicate the mutations in ARGs nor tell the existence of the remaining sequences outside of this range. This disadvantage makes qPCR easy to generate false positive results. Another drawback of qPCR is the limited number of targeted genes. Multiplexed qPCR and chip-based high-throughput qPCR microarrays are applied to increase the number of targets in one qPCR run. Multiplexed qPCR can target up to 20 genes in one reaction (Jacky et al., 2021), but the interaction among different primers and probes may become complex when the number of targets increases and make the results inaccurate (Xie et al., 2022). Chip-based high-throughput qPCR microarrays can run thousands of qPCR reactions in the meantime by separating different assays into droplets, allowing it to target dozens to thousands of genes (Fassy et al., 2021). This method can avoid the disadvantage of complex interactions of different primers and probes in conventional multiplexed qPCR while largely increasing the number of targets. However, to ensure good amplification efficiencies, the design and validation of specific primers for each individual ARG target is required. When the number of targets increases, the cost of primer design and validation will also increase. High-throughput qPCR thus has a hidden cost in the large number of primer validation. In addition, not all chip-based high-throughput qPCR machines can adjust the numbers of samples and targets flexibly (Lee et al., 2023). Some chip-based high-throughput qPCR machines have upper limits of 96 or 192 genes to be targeted in one run (Olwagen et al., 2019). In contrast, metagenomics does not require primer design for specific targets, so that it can detect a wide range of ARGs without pre-estimation (de Abreu et al., 2020). Next-generation sequencing (NGS) is increasingly accessible nowadays with the development of cost-friendly benchtop sequences. However, because of the low percentage of ARGs in wastewater DNA samples, most of the reads obtained in conventional metagenomics are not relevant to ARG detection. The sensitivity of metagenomics is lower than qPCR, making the detection of low-abundance ARGs difficult by metagenomics (Liguri et al., 2022).

To increase the sensitivity for ARG detection using metagenomics, different methods have been developed. Hybridization capture is one of the commercialized multiplexed target-enrichment methods for NGS. This method adds an additional step after indexed library preparation using streptavidin magnetic beads and specifically designed biotinylated DNA probes to capture DNA fragments with matched sequences (Guitor et al., 2020). Thousands to millions of probes can be used to target various DNA sequences (Samorodnitsky et

al., 2015). However, this method is expensive, complex, and time-consuming because of the synthesis of highly multiplexed biotinylated DNA/RNA probes (Sundararaman et al., 2023, Gaudin and Desnues, 2018). Recently, CRISPR (clustered regularly interspaced short palindromic repeats-Cas9-based target enrichment sequencing methods were developed for clinical diagnosis (Quan et al., 2019, Gilpatrick et al., 2020). Such methods use specifically designed guide RNA (gRNA) to guide Cas9 nuclease specifically leave the target genes to generate 5' phosphate group for adapter ligation. The libraries generated by this method can be applied to both the Illumina NGS platform and the Nanopore third-generation sequencing platform (Quan et al., 2019, Gilpatrick et al., 2020). Unlike qPCR where multiplexed primers and probes may have complex and unexpected interactions (Xie et al., 2022), the interaction of Cas9-gRNA complexes to DNA samples is simpler, making the Cas9 cleavage reactions possible to be highly multiplexed (McCarty et al., 2020). Unlike biotinylated DNA probes in hybridization capture (which do not contain conserved regions), the gRNA used for Cas9 cleavage contains conserved regions (Liang et al., 2015), making the synthesis of multiplexed assays of CRISPR-Cas9 enriched methods more flexible and economical.

NGS is increasingly accessible to laboratories because of its reduced cost (McCombie and McPherson, 2019). In this study, we developed a novel CRISPR-NGS method for ARG detection in wastewater samples. The detailed step-by-step protocol for the entire CRISPR-NGS workflow can be found at protocols.io (Mao and Nguyen, 2023). This CRISPR-NGS method includes six major steps (Fig. 1). Step 1 is to prepare a highly multiplexed gRNA pool (i.e., 6010 different sequences) targeting different sites in ARGs specifically. We chose to synthesize trans-activating CRISPR RNA (tracrRNA) and multiplexed CRISPR RNA (crRNA) from DNA templates to reduce the cost because purchasing gRNA individually will be extremely costly (Liang et al., 2015). Step 2 is to prepare DNA samples for sequencing. This step includes three sub-steps of wastewater DNA extraction, DNA blocking by 5' phosphate group removal, and DNA positive control spike. In NGS library preparation, adapter ligation is a necessary step which requires DNA fragments have 5' phosphate group. To avoid adapters ligating to DNA fragments irrelevant to ARG detection, we used alkaline phosphatase to remove 5' phosphate groups from all DNA fragments before library preparation. After blocking, a 250-bp DNA fragment with synthesized 5' phosphate group obtained from Pseudogubenkiania sp. strain NH8B which should be absent in most DNA samples was spiked in the DNA sample as the positive control for sequencing (Zhang and Ishii, 2018). Step 3 is to specifically cleave the DNA samples at the sites in the ARGs using Cas9 nuclease to generate new 5' phosphate groups for adapter ligation. The newly generated 5' phosphorylation ARG fragments were further modified by a dA-tailing step for optimized adapter ligation efficiency. Step 4 is to ligate the adapters to the specifically cleaved ARG fragments. Step 5 is to amplify the adapter-ligated fragments by polymerase chain reaction (PCR) using the primers targeting the conserved region on the adapters to enrich the ARG fragments in the DNA sample. Step 6 is to check the fragment lengths and the DNA concentrations of the library and load the libraries for NGS. As described here, the CRISPR-NGS method modifies the DNA fragmentation step in NGS library preparation, so that the sensitivity of NGS in ARG detection can be largely increased. With the same total number of reads, hundreds more ARGs in low abundances can sewage samples can be uncovered, and a lower detection limit can be expanded from the magnitude of 10^{-4} to 10^{-5} as determined by qPCR relative abundance.

2. Methods

2.1. Wastewater sample collection, pre-treatment, and DNA extraction

The wastewater samples used in this study included five influents from the wastewater treatment plants in Chicago, Illinois, and one raw sewage composite samples directly collected from a manhole. More information on sample collection and processing is in Supplementary Methods S1.

2.2. Mock community DNA sample preparation

We estimated the false positive and false negative detection of ARGs using the mixture of DNA extracts at equal concentrations from 10 bacteria isolates (three E. coli from wastewater, three E. coli from clinical sources, one Shigella spp. from a clinical source, two Salmonella spp. from clinical sources, and one Pseudomonas aeruginosa PA14 (see Table S1) (Grace et al., 2022). This mock community was subjected to both whole-genome sequencing and CRISPR-NGS to determine the false positive and false negative rates of the

CRISPR-NGS method. Details on DNA extraction, preparation, and subsequent whole-genome sequencing are in Supplementary Methods S2. The whole-genome sequences are available in BioProject PRJNA1147862.

2.3. Design and synthesis of crRNA and tracrRNA

Multiplexed gRNA for ARG detection was designed by FLASHit

Fig. 1. The workflow of the CRISPR-NGS method.

(Quan et al., 2019). Four thousand six hundred and forty-one ARG nucleotide sequences for gRNA design input were downloaded from the CARD protein homolog model on September 12, 2022 (Alcock et al., 2020). The default exclusion of off-targets in the E. coli BL21 genome in FLASHit was removed, because the Cas9 protein used in this study was commercially purchased. In total 6010 different 20-nt spacer sequences were generated by FLASHit to cleave target ARGs into approximately 200-bp fragments. More details of the synthesis of crRNA and tracrRNA are in Supplementary Methods S3.

2.4. CRISPR-NGS and regular NGS library preparation and sequencing

We used Cas9 nuclease to specifically cleave targeted sites for DNA fragmentation in CRISPR-NGS library preparation. For regular NGS library preparation, we used non-specific enzymes for DNA fragmentation. A synthesized 250-bp DNA fragment obtained from Pseudogulbenkiania sp. strain NH8B with 5' phosphate group was spiked to all libraries as the positive control (Zhang and Ishii, 2018). In the adapter ligation step, xGen UDI-UMI Adapters (Integrated DNA Technologies) was used to reduce the PCR bias. Detailed information for CRISPR-NGS library preparation, regular NGS library preparation, library quality control, and Illumina sequencing is in Supplementary Methods S4-S5.

2.5. Sequencing read mapping

After paired-end sequencing, UMI-tools, PriceSeqFilter, Bowtie2, and SAMtools were used for read deuplication, quality control, and ARG mapping (Smith et al., 2017, Ruby et al., 2013, Langmead and Salzberg, 2012, Danecek et al., 2021, Li et al., 2009). The same read mapping methods were used for both CRISPR-NGS and regular NGS libraries. All sequencing raw data, including the separated UMI index files, can be found in BioProject PRJNA1148079. The reference file for read mapping includes the ARG sequences downloaded from CARD (see 2.3) and the sequence of the 250-bp NHBB DNA positive control. Any AGs with coverages higher than 75% were considered present in the sample, according to previous studies (Chen et al., 2016, Fang et al., 2015, Li et al., 2020). The identity threshold was approximately 90%, according to the end-to-end alignment algorithm of Bowtie2 (Langmead and Salzberg, 2012). Detailed information for read mapping is in Supplementary Methods 56.

3. Results

3.1. Method validation

We validated that alkaline phosphatase blocking inhibited the T4 DNA ligase from connecting different DNA fragments together. The detailed method for the validation of 5' phosphate group removal can be found in Supplementary Methods S7. After phosphatase blocking, the 5' phosphate group of DNA fragments is removed, and T4 ligase could not anneal DNA fragments to form new fragments of other sizes (Fig. S1). We used a synthetic 731-bp DNA template with 5' phosphorylation to test the blocking efficiency of alkaline phosphatase. After phosphatase blocking, the DNA was incubated with T4 ligase. As shown in Fig. S1, while the DNA sample without blocking showed multiple bands, the blocked DNA remained its original 731-bp size, suggesting the 5' phosphate groups of blocked DNA fragments were removed, and T4 DNA ligase could not ligate these DNA fragments.

We validated the Cas9 cleavage using an in-silico analysis that compares the breakpoints in raw sequencing reads with the intended Cas9 cleavage sites. In read mapping, paired 150-bp raw sequencing reads were aligned to the referenceARG sequence obtained from CARD. The paired reads, which started from different ends of the same DNA fragment, should have an overlapping region if the DNA fragment is smaller than 300 bp. If the reads are from different fragments, there would be no overlap. Because most DNA fragments were designed to be around 200 bp using a protocol developed previously (Quan et al., 2019) and the maximum

read length was 150 bp in NGS, the entire DNA fragment would not be covered by one single read but could be covered by two paired reads from different directions. Therefore, in the read mapping plot of a cleaved 200-bp DNA fragment, we expected to see the middle part (usually around 100 bp) overlapped by the paired 150-bp reads from different directions and the remaining left and right part covered by the remaining lengths of the paired reads. Because the DNA fragments were generated by Cas9 cleavage at a few designed sites, these DNA fragments should not overlap with each other in the read mapping of the ARG. The locations where two DNA fragments meet each other should be the same as the designed Cas9 cleavage sites. The example of overlapping regions on different DNA fragments in the same ARG was marked and shown in Fig. S2, which shows the read mapping plot for CRISPR-NGS and regular NGS for the same gene OXA-36 (ARO: 3001430) (Alcock et al., 2023). The four cleavage sites on OXA-36 broke the read mapping histogram of CRISPR-NGS into three overlapping regions while the regular NGS read mapping histogram was smooth. We checked the read mapping of 79 ARGs from different ARG families, 78 of the raw reads were broken at the designed cleavage sites except for ropB

which contained a few off-target cleavage sites. This finding suggested most of the ARGs detected by CRISPR-NGS truly existed in the samples.

3.2. CRISPR-NGS method showed high accuracy in ARG detection in the mock community sample

We conducted whole-genome sequencing of 10 bacteria isolates (listed in Table S1) and these genomes served as a basis to determine false positives and false negates in ARG detection by the CRISPR-NGS method. Because some ARGs share highly similar sequences, we divided the 4641 targeted ARGs into 1208 clusters based on their similarities in terms of ANI to avoid overcounting these ARGs (Fig. S3; Spreadsheet S1). The details of ARG clustering method are in Supplementary Methods S8. We first used annotation for the 10 whole-genome sequences to find 100 ARGs in 99 clusters existing in the mock community. The CRISPR-NGS method detected 1032 ARGs in 124 clusters in the same sample (Spreadsheet S2). Ninety-seven clusters were found by both CRISPR-NGS and annotation of the whole genomes. Because the algorithm of whole-genome annotation may miss some ARGs (Maboni et al., 2022), we used BLAST to search the ARG sequences listed in CARD within the 10 whole bacterial genomes in the mock community (Madden, 2013). We found 25 additional clusters of ARGs that were detected by the CRISPR-NGS method in the mock community. Finally, we used Sanger sequencing to confirm the presence of two clusters of ARGs that were not found by either annotation or BLAST in the 10 whole-genome sequences, one containing quinolone resistance protein (qnr) genes and the other containing one single ARG aph(3)-Ia. The primer sequences of these two genes are listed in Table S2. The amplicon of the qnr85 gene was aligned to the CARD reference sequence in 99.65% identity, suggesting this gene was missed by long-read PacBio sequencing but could be detected by short-read NGS. In total, as summarized in Table S3, we confirmed the presence of 123 ARGs cluster detected by the CRISPR-NGS method (97 from annotation as shown in Spreadsheet S3, 25 from the BLAST method, and one from Sanger sequencing. These clusters are the true positives. There were two false negatives, which were found by whole-genome sequencing but not by CRISPR-NGS. The maximum identities of these two ARG clusters in all mock community genomes were only 86–87% when comparing with the reference sequences, while the maximum identities of any other true positive ARGs were equal to or greater than 93%. The NGS libraries used Bowtie2 for ARG identification which has an identity threshold of approximately 90% under our current settings (Langmead and Salzberg, 2012). In contrast, ARGs were identified from the whole-genome sequences by the Genome Animation tool on Bacterial and Viral Bioinformatics Resource Center (BV-BRC) with an identity threshold of 80% (Alcock et al., 2020, Kent, 2002, Olson et al., 2023). The difference in identity thresholds made the two ARG clusters of 86–87% identities missed by NGS.

We also compared the ARG detection results for the mock community sample to the regular NGS method using the similar methods as described above. The regular NGS library detected two more ARG clusters, dfrA14 and adA13, than the CRISPR-NGS library (Spread-sheet S4). However, the identities of these two ARGs were 89% in BLAST against the reference sequence. The coverage of adA13 (ARO: 3002613, 799 bp) in the CRISPR-NGS

library was 65.91%, slightly below the 75% threshold. It is worth noting that regular NGS didn't have any reads mapped to the reference sequence after approximately 700 bp, suggesting potential accumulation of mutations there (Fig. S4). The cleavage site of CRISPR-NGS at 772-795 bp might be affected due to the mutations and resulted in the low coverage. The failure in dfrA14 (ARO: 3002859, 475 bp) detection might be due to the unsuccessful cleavage at 15-38 bp (Fig. S4). To conclude, though CRISPR-NGS may miss some ARGs with accumulated mutations that make the identities of the ARGs below 90% with reference sequences, its overall accuracy in ARG detection is high.

3.3. High reproducibility in ARG detection was observed in the CRISPR-NGS libraries made from the same DNA sample

The reproducibility of the CRISPR-NGS method was tested by comparing the results of two sewage samples collected from Calumet on 8/16/2023 and 10/24/2023. The detailed method is in Supplementary Methods S9. The DNA extracted from these samples were made into duplicated libraries. Nine hundred and sixty ARGs were detected by both libraries for Calumet (8/16/2023), with 1020 and 1045 ARGs detected in each library (94.1% and 91.9% were detected in both). As for Calumet (10/24/2023), 1597 and 1281 ARGs were detected in each duplicate, while 1263 ARGs were shared in common (79.1% and 98.6%) were detected in both).

The read depths for the ARGs detected in only one duplicated library ranged from 1.03 to 291.68 with a median of 7.70 for Calumet (8/16/2023) and ranged from 1.08 to 1041.34 with a median of 20.72 for Calumet (10/24/2023). These ARGs were missed because their coverages were below the 75% threshold, instead of no read alignment. For example, the 987-bp AAC6,30,AAC6,Ib gene (ARO: 3002599) detected only in one of the two libraries of Calumet (10/24/2023) with read depth of 1041.34 and coverage of 82.07%, while the other library only got a coverage of 60.59%. As a comparison, we also made duplicated regular NGS libraries for Calumet (10/24/2023). One hundred and fifty-six ARGs were detected by both libraries, while 399 and 190 ARGs were detected by each individual library (39.1% and 82.1%). The read depths of the ARGs detected only by one library ranged from 0.91 to 6.07, with a median of 1.36. These findings suggest that though having 100% same ARG detection by metagenomics is difficult, it is still possible to repeatedly detect most of the ARGs in the same DNA sample by CRISPR-NGS.

Then we compared the read depths of the ARGs detected in both duplicated libraries. The read depths of shared ARGs were fitted linearly in log-10 scale with slopes of 0.92 and 1.05, intercepts of 0.017 and -0.377 (Fig. 2). The R^2 values were 0.90 and 0.98, respectively, suggesting low fitting variations. The slopes of both fittings were close to 1, and the intercepts were close to 0, suggesting the read depths of the ARGs detected in the duplicated libraries were reproducible.

3.4. More ARGs were detected by the CRISPR-NGS method than regular NGS method

The improved sensitivity of the CRISPR-NGS method in ARG detection was evaluated by comparing the number of ARGs and ARG families detected in the same sewage sample by the CRISPR-NGS method and regular NGS method, respectively. Fig. 3 shows the number of ARGs and ARG families that were mapped with over 75% coverage by only the CRISPR-NGS method, only the regular NGS method, and both methods. The sewage samples had 288-433 ARGs detected by regular NGS libraries, and these numbers increased to 542-1615 for CRISPR-NGS libraries. Only from seven to 43 ARGs were uniquely detected by regular NGS libraries, while 278-1189 ARGs were uniquely detected by CRISPR-NGS libraries. The ARGs detected in wastewater samples by different CRISPR-NGS and regular NGS libraries are listed in Spreadsheet S5 and S6. These additional ARGs detected by CRISPR-NGS libraries were mostly from ARG families that were not covered by regular NGS libraries. Seventeen to 61 more ARG families were only detected by CRISPR-NGS libraries for the sewage samples, while zero to three were only detected by regular NGS libraries. CRISPR-NGS libraries missed these ARG families mainly because the coverages of these ARGs were slightly lower than the 75% threshold. For example, CfxA beta-lactamase genes were missed by three CRISPR-NGS libraries (Calumet 10/24/2023, Calumet 12/19/2023, and Lemont 12/19/2023). The coverages of CfxA2 (ARO: 3003002) in the 72.05% and 72.57%, less than 3% below 2 below. The ARG families detected only by CRISPR-NGS libraries.

Fig. 2. The ARG read depths in duplicated libraries made from the same DNA samples (a: Calumet, 8/16/2023; b: Calumet, 10/24/2023) fitted linearly in log-10 scale with R^2 values of 0.90 and 0.98, respectively.
Fig. 3. More ARGs (a) and ARG families (b) were detected by CRISPR-NGS method than regular NGS method in different sewage samples. The yellow bars show the number of ARGs and ARG families only detected by CRISPR-NGS method, the purple bars show the number of ARGs and ARG families detected by both CRISPR-NGS and regular NGS methods, and the dark blue bars show the number of ARGs and ARG families only detected by the regular NGS method. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
example is the mobilized colistin resistance (MCR) phosphoethanolamine transferase genes that encode resistance to colistin which is one of the last-resort antibiotics (Han et al., 2020, McKenna, 2013). The CRISPR-NGS method detected MCR genes in a high diversity, including MCR-3, MCR-4, MCR-5, MCR-9, and MCR-10, while regular NGS only detected MCR-3 or MCR-5 in some libraries.

3.5. CRISPR-NGS method could detect ARGs in higher read depths than regular NGS method

The ARGs detected by CRISPR-NGS libraries also had higher read depths than those detected by regular NGS libraries. The fold changes in read depths of the 145 ARGs detected in all sewage samples by both CRISPR and regular NGS libraries are plotted in Fig. 4. All ARGs had significantly higher read depths in CRISPR-NGS libraries than regular NGS libraries. The fold changes in read depths varied from 2.73 to 126.45, suggesting the CRISPR-NGS method can detect ARGs in higher confidence than regular NGS method.
To compare the ARG detection efficiencies of CRISPR-NGS and regular NGS with deeper sequencing read depths, we subsampled the CRISPR-NGS reads and compared the ARG mapping results with the corresponding regular NGS libraries without subsampling (Fig. 5). The subsampling protocol can be found in Supplementary Methods S10. Before subsampling, the total number of reads for CRISPR-NGS libraries for sewage samples ranged from 43.3 million to 58.2 million, and the total number of reads for regular NGS libraries ranged from 44.6 million
Fig. 4. The read depths of all 145 ARGs detected in all sewage samples were significantly higher in CRISPR-NGS libraries than regular NGS libraries. The y axis shows the p-values in a reversed log-10 manner, and the x axis shows the fold changes in read depths (CRISPR-NGS versus regular NGS). The horizontal dash line means the threshold of a significant p-value equals to 0.05, and the vertical dash line means the threshold in fold change of 1.5.
to 60.7 million (Table S4). When subsampling 2% of the reads, three CRISPR-NGS libraries (Calumet 12/19/2023, Calumet 10/24/2023 A and B) detected more ARGs than the corresponding regular NGS libraries. When the subsampling ratio reached 10%, all CRISPR-NGS libraries except for Lemont (12/19/2023) detected more ARGs than corresponding regular NGS libraries. The CRISPR-NGS library for

Fig. 5. When the total read depth reaches approximately 2–20% of current read depth, CRISPR-NGS libraries can detect similar numbers of ARGs as regular NGS libraries with their original read depths. The CRISPR-NGS libraries were subsampled by ratios of 1%, 2%, 5%, 10%, 20%, 50%, and 70%. The horizontal dash line at y=1 means equal number of ARGs detected by the subsampled CRISPR-NGS library and the regular NGS library with original read depth.
Lemont (12/19/2023) detected 89.6% of the number of ARGs detected by its regular NGS library when doing 10% subsampling, and the percentage increased to 99.2% at subsampling ratio of 20%, and 171.2% at subsampling ratio of 50%. At subsampling ratio of 50%, all curves reached a plateau phase (Fig. 5). These findings suggest that CRISPR-NGS can detect similar numbers of ARGs as regular NGS with approximately 2–20% of the total number of reads. With approximately 25 million total number of reads (i.e., approximately 50% of the read depths used in this study), the number of ARGs detected by CRISPR-NGS approaches maximum. This improvement suggests that CRISPR-NGS can make it possible to detect ARGs deeply by metagenomics on many benchtop sequencers that can generate more than 25 million reads in one run, instead of ultra-deep sequencing for regular NGS libraries that usually require expensive and professional

production-scale sequencers.

3.6. ARG presence could be validated by Sanger sequencing.

We used Sanger sequencing to validate the presence of selected ARGs in sewage samples (Table 1). The ARG targets were amplified by PCR, and the PCR amplicons were Sanger sequenced and compared to the reference sequences in CARD. The detailed method for Sanger sequencing validation can be found in Supplementary Methods S11.

Table 1

The Sanger sequencing identities and adjusted coverages for the ARGs detected in different CRISPR-NGS libraries.

Gene	Identity (%)	Coverage adjusted by amplicon lengths (%)
aph(3')-Ib	99-100	95-98
sul1	97-99	95-96
CRP	89-94	95
sul2	95-99	92-96
aph(3')-Ia	95-99	73-97
dfrA1	96-97	93-95
GES-26	95-100	96-97
MCR-5.1	98	77-99
MCR-4.1	98	98
TEM-1	99	96
ac(3)-IId	96-99	58-93
tet(G)	94	88
qnrB5	100	97

Note: The coverages were adjusted by dividing the aligned length by the amplicon length, rather than the coverage provided in BLAST, which divides the aligned length by the query length (i.e., the reference ARG length in CARD).

We chose Sanger sequencing for validation, because Sanger sequencing for PCR amplicons is accurate, cost-effective, and fast when there are only a few targets. The identities between the sequenced PCR amplicons and the reference sequence of the ARGs in CARD were between 89% and 100%, suggesting the amplicon sequences were highly similar to the reference sequences. Because the PCR primers designed for different ARGs had different binding regions, the amplicons could not cover the full lengths of the ARGs. After adjusting with the amplicon lengths, the coverage of the ARGs were between 58% and 99%. The 58% coverage of aac(3)-IId and 73% coverage of aph(3)-Ia were due to the low quality of Sanger sequencing of the Campus Town manhole sample. After removing these outliers, the coverages were between 77% and 99%. The high coverages and identities of Sanger sequencing also matched the read mapping plots. For example, the two colistin resistance genes, MCR-5.1 and MCR-4.1, were nearly fully mapped by CRISPR-NGS reads (Fig. S5), though these two genes were barely detected by regular NGS.

3.7. CRISPR-NGS libraries had lower detection limit than regular NGS libraries, and the read depths in CRISPR-NGS libraries were proportional to the relative abundances determined by qPCR

We evaluated the detection limit of the CRISPR-NGS method by plotting the read depths against the relative abundance detected by qPCR for nine ARGs with various read depths detected by CRISPR-NGS or regular NGS methods in sewage samples (Fig. 6). The detailed method for qPCR can be found in Supplementary Methods S12. The ARGs and corresponding qPCR primers are listed in Table S5. The Cq values of the 16S rRNA gene in the sewage samples are listed in Table S6. The read depths in both CRISPR-NGS and regular NGS libraries increased when the qPCR relative abundances increased, suggesting the read depths of CRISPR-NGS and regular NGS can be used for ARG abundance estimation. As shown in Fig. 6, the lower detection limit of the CRISPR-NGS method for these nine ARGs was down to 2.23×10^{-5} copies ARG per copy of 16S rRNA gene, while the lower detection limit of regular NGS method was at 1.79×10^{-4} copies ARG per copy of 16S rRNA gene. Thus, the CRISPR-NGS method has a higher sensitivity than regular NGS method in argG detection.

A previous study reported varied relative abundances of ARGs in wastewater samples ranging from 1.4 \times 10^{-5 to 0.09 (Srathongneam et al., 2024), suggesting the improved lower detection limit of CRISPR-NGS can promote the detection of ARGs in low abundances in wastewater samples.

Fig. 6. The read depths versus the qPCR relative abundances of the nine ARGs detected in different wastewater samples. The orange dots are the read depths of the ARGs detected in the CRISPR-NGS libraries. The blue dots are the read depths of the ARGs detected in the regular NGS libraries. The vertical dash lines in corresponding colors indicate the lower detection limits of CRISPR-NGS and regular NGS in terms of qPCR relative abundance. CRISPR-NGS had a lower detection limit of 2.23 \times 10^{-5, as labeled by the orange vertical line, while the lower detection limit of regular NGS was 1.79 \times 10^{-4, as labeled by the blue vertical line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.8. CRISPR-NGS is better at detecting low abundance ARG than regular NGS

The most abundant clusters of ARGs (i.e., ARG clusters with read depths greater than 200 in any of the CRISPR-NGS libraries) in the sewage samples were selected and plotted into heatmaps for comparison in Fig. 7. These 41 ARG clusters were distributed in eight major drug classes, including aminoglycoside (8), beta-lactam (17), antisptics (2), fluoropinolone (2), macroline (3), multi-drug (3), phenicol (2), and tetracycline (4). The ARG clusters in Fig. 7 are named based on the ARG names and families in the cluster. The multi-drug class was defined to include the ARGs that are resistant to at least two drug classes in the 61 sub-terms under the CARD “antibiotic molecule” ontology term (Alcock et al., 2023). Comparing between CRISPR-NGS and regular NGS libraries, the entire group of CMY-2 and KPC beta-lactamase genes and qnrB8-like quinolone resistance genes were missed by regular NGS. Comparing the wastewater treatment plant sewage samples, the campus town manhole sample had lower abundances of GES, MOX-9, and OXA-427-like beta-lactamase genes than the wastewater treatment plant (WWTP) influent samples. The abundance of the tet(C) gene in the Campus Town manhole sample was greater than any of the Chicago WWTP influent samples, but tet(E) in the Campus Town manhole sample was lower than the WWTP influent samples. The above findings suggested different ARG distributions in different geographical areas or different sampling scales. As for the ARGs shared in common, sul1, qacE/qacDelta1, qacL, and OXA-2-like beta-lactamase genes were constantly more abundant than the other ARGs in all sewage samples. Since sul1, qacE, and qacEdelta1 are known to be located on class 1 integrons which play an important role in the spread of ARGs, our findings suggest that class 1 integron may be actively transferring ARGs in these wastewater samples (Ghaly et al., 2017).

When extending the scale to all ARGs with high or low abundances, the profiles detected by CRISPR-NGS and regular NGS were more different (Fig. S6). The entire class of aminocoumarin antibiotics was not detected by regular NGS libraries but by CRISPR-NGS libraries. Fewer classes of beta-lactamase genes were detected by regular NGS libraries than CRISPR-NGS libraries. For example, KPC and CTX-M were not detected by any regular NGS libraries but by CRISPR-NGS libraries. Similarly, regular NGS libraries missed several clusters of multi-drug, peptide, diaminopyrimidine, and phenicol resistance genes which were captured by CRISPR-NGS libraries. But for aminoglycoside, macroide, and tetracycline resistance genes, regular NGS libraries detected similar numbers as CRISPR-NGS libraries. Such findings suggest that when the range of ARGs for surveillance expands, CRISPR-NGS works better in capturing the ARGs in the classes that are less abundant than the three drug classes mentioned above.

4. Discussion

Metagenomics has been applied widely in ARG surveillance in different water and wastewater samples (Garner et al., 2018, Yin et al., 2019, Li et al., 2022, Zhou et al., 2021, Davis et al., 2023). The advantage of metagenomics is that it can detect a wide range of ARGs without knowing the targets. However, it is challenging to sequence ARGs present in low abundances by metagenomics. We overcome this disadvantage by using CRISPR-Cas9 with multiplexed gRNA to specifically enrich the ARGs in wastewater DNA samples during

NGS library preparation.
After the CRISPR-Cas9 treatment, compared to regular NGS libraries, up to 1189 more ARGs were detected, and CRISPR-NGS libraries only
Fig. 7. The heatmaps for the most abundant ARGs (i.e., the ARGs with read depths greater than 200 in at least one library detected by CRISPR-NGS) in different sewage samples; a: The read depths of the most abundant ARGs detected by CRISPR-NGS libraries; b: The read depths of the most abundant ARGs detected by regular NGS libraries. The read depths increase from green to pink. The white color means the ARGs was not detected in the sample by the corresponding method. From top to bottom, the heatmaps were divided into eight parts by horizontal black lines according to the drug classes of the corresponding ARGs: aminoglycoside, beta-lactam, antiseptic, fluorquinolone, macrolide, multi-drug, phenicol, and tetracycline. The solid vertical lines divide samples by different locations of collection, and the dash vertical lines divide samples from the same location by different collection times. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

require 2–20% total number of reads to detect similar number of ARGs as regular NGS libraries. Based on the findings in this study, CRISPRNGS works the best for samples that contain highly diverse ARGs in relatively low abundances. Various types of extended-spectrum beta-lactamase (ESBL) genes were detected in the sewage samples in this study by CRISPR-NGS. These genes include not only the common ones like TEM, CTX-M, and OXA, but also the rare ones, such as GES, PER, VEB, BES, and TLA (Castanheira et al., 2021). Among these genes, GES, PER, OXA, and VEB were abundant enough to be constantly detected by regular NGS. Other genes were too diluted to be stably detected without CRISPR treatment before NGS, including the clinically important carbapenem resistance gene KPC and cefotaxime resistance gene CTX-M. The return is low but extra effort is required to monitor these rare ESBLs using conventional methods because their abundances were too low to be detected in most cases. However, studies showed that these ESBLs were also evolving and spreading globally (Pagani et al., 2004, Gniadkowski, 2008). For example, GES-2 gained resistance to carbapenem by only a single point mutation (Poirel et al., 2001). Though rarely detected in the clinical isolates obtained until now, our findings indicated that these ESBL genes were circulating in human communities and could potentially cause public health problems in the future. For example, though considered as a less common ESBL, the read depths of GES in WWTP samples were in the top 17 of all 76 detectable beta-lactamase gene clusters. In contrast, GES was not as abundant as in WWTP samples for the Campus Town manhole sample. This finding suggested that the presence and abundance of different ESBL may be dynamic. Innovative high-throughput detection methods with high sensitivity such as CRISPR-NGS will be helpful in the surveillance of the emergence of such unexpected ARGs.
The CRISPR-NGS method is promising for ARG detection, especially for the surveillance of sewage samples which may have many unexpected targets. Early warning of new threats in public health is challenging, because lots of new targets may be hidden in low abundance in human communities and suddenly emerge to cause an outbreak (diamond et al., 2022). Surveillance methods should be high in throughput and sensitivity simultaneously. When the throughput is high, multiplexed or chip-based qPCR is labor-intensive in primer design and sample preparation. Because of such reasons, NGS is more suitable for monitoring unexpected targets than qPCR. Conventional NGS can increase its sensitivity by increasing the total number of reads during a sequencing run (Zaheer et al., 2018, Gweon et al., 2019). However, this method can be costly, because increasing the total read depth by 10-fold means the total sequencing cost for the same sample is increased by 10-fold. Currently, the Illumina sequencing cost is around 2.2 per million reads (Pricing | Biotech). For a library with 50 million reads, the cost solely for sequencing is around 110. If the total number of reads increases by 10-fold to 500 million, the sequencing cost for the same library will increase to around 1100. Considering more than 99.9% reads are irrelevant to ARG detection (Guitor et al., 2020), increasing the proportion of ARG reads before loading the samples to sequencing machine will

be more economical. The CRISPR-NGS method could detect the same number of ARGs read depth using 2–20% of total number of reads required by regular NGS libraries with only an increased cost of around 30 in Cas9 nuclease and gRNA preparation for each library. The relatively low cost makes the CRISPR-NGS method attractive for routine ARG surveillance in the future. The reduced read depth requirement of CRISPR-NGS make it possible to use downgraded but less expensive benchtop sequencers like MiSeq to detect ARGs by metagenomic sequencing and generate results comparable to the more advanced production-scale sequencers like NovaSeq.

Although third-generation sequencing can help identify the hosts and mobile genetic elements by analyzing the long DNA fragments upstream and downstream of ARGs that are generated as the byproduct during Cas9 cleavage (Fuhrmeister et al., 2024), the limitation of combining CRISPR with third-generation sequencing is that the enrichment of ARGs will be less efficient than NGS, because the long DNA fragments are difficult to be amplified to sufficient concentrations for sequencing. Due to this limitation, the required DNA input for CRISPR-enriched third-generation sequencing is 1–10 g (Gilpatrick et al., 2020). As a comparison, CRISPR-NGS only requires 100–300 ng DNA input. The combination of CRISPR and third-generation sequencing will work for bacterial isolates and fecal samples that have high DNA concentrations, as previously reported (Fuhrmister et al., 2024). Wastewater DNA is usually more diluted which is not sufficient for Cas9-enriched third-generation sequencing. To apply CRISPR-based target enrichment to third-generation sequencing, a better amplification method and a modified gRNA pool for generating long DNA fragments in kilobases need to be developed in future studies.

There are also limitations of the CRISPR-NGS method for future improvement. First, genetic information other than ARGs, for example, the pathogen genomes and mobile genetic elements, may be missed by CRISPR-NGS method, because these sequences are not targeted during assay design. ARGs that are currently unknown may also be missed for the same reason. Identifying ARG hosts can help find ARG-carrying pathogens and guide clinical drug use. To detect these targets, future studies can include the sequences of genetic markers of pathogens and mobile genetic elements in gRNA design. Second, the read depths of ARGs may be biased because of the high number of PCR cycles used during CRISPR-NGS library preparation if using conventional adapters. UDI-UMI adapters with molecule-specific adapter sequences are recommended to reduce the PCR bias, because the over-amplified DNA fragments can be identified by their shared UMI indexes. Third, CRISPR-NGS cannot quantify ARGs in absolute copy numbers as qPCR, the relative quantification is also not as accurate as qPCR. We recommend using CRISPR-NGS periodically for wastewater surveillance to identify ARGs relevant for public health and use qPCR to quantify these ARGs in a more frequent manner. Fourth, between seven to 43 ARGs were missed by CRISPR-NGS but captured by regular NGS in sewage samples, suggesting not all ARGs were successfully targeted by CRISPR-NGS. This may be due to the low on-target cleavage efficiency of some gRNA or due to “dropouts” in template DNA oligo pool synthesis. Though Cas9 nuclease can cleave DNA at targeted sites with a few mismatches (Boyle et al., 2021), the efficiency of Cas9 binding and cleavage is reduced when mismatches occur. In addition, the current DNA oligo pool synthesis technique cannot guarantee 100% synthesis of all oligos. Around 0.8% of the sequences in a DNA oligo pool can be missed during synthesis. For the oligo pool used in this study with 6010 different sequences, the number of “dropout” oligos is around 48. The corresponding gRNA would not be transcribed and used in library preparation. Selecting a vendor with relatively lower “dropout” ratio for oligo pool synthesis may help increase the on-target cleavage ratio. Fifth, the CRISPR-NGS method may require a more flexible read mapping threshold (e.g., different coverage threshold for different ARGs) to avoid false negative results, making the bioinformatic pipeline after sequencing more complex. In this study, we used a fixed coverage threshold of 75% to make CRISPR-NGS results comparable to regular NGS, but future studies can develop a bioinformatic pipeline with flexible threshold specifically for CRISPR-NGS. Because of the recognition of the protospacer adjacent motif (PAM) regions, gRNA sequences cannot be designed exactly at the beginning and the end of the ARGs. This disadvantage in gRNA design can result in reduced coverage in read mapping of some ARGs, causing false negative results of these ARGs when using a fixed threshold.

5. Conclusions

Here, we introduced a CRISPR-Cas9 modified NGS library preparation method to increase the sensitivity in ARG detection in sewage samples. Up to 61 ARG families or up to 1189 ARGs present in low abundance were uncovered by NGS after CRISPR-Cas9 modification. With the same total read depth for one sequencing run, CRISPR-NGS libraries lower the detection limit from magnitude of 10^{-4 to 10^{-5 as

determined by qPCR relative abundance. CRISPR-NGS could detect similar number of ARGs as regular NGS with only 2–20% required total number of reads. CRISPR-NGS only requires approximately 25 million reads to detect ARGs deeply in sewage samples, making it possible to be applied to cost-friendly and more accessible benchtop sequencers.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yuqing Mao reports financial support was provided by Water Research Foundation. Yuqing Mao reports financial support was provided by NTU/U of I Joint Research and Innovation Seed Grants Program. Yuqing Mao reports financial support was provided by United States Environmental Protection Agency. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We acknowledge funding from the Water Research Foundation (5182), NTU/U of I Joint Research and Innovation Seed Grants Program, and USEPA (R840487). This study has not been formally reviewed by EPA. EPA does not endorse any products or commercial services mentioned in this publication. We acknowledge the support from Dr. Xing Wang, Lucas D. Akin, and Abhisek Dwivedy. We acknowledge the effort in sample collection from Arthur R. Schmidt IV, Ryan McAllister, Amaja Craft, and Josie Hoppenworth. We acknowledge Dr. Rachel J. Whitaker, Dr. Jayadevi H. Chandrashekhar, and Laura C. Suttenfield in sharing Pseudomonas aeruginosa PA14. We acknowledge Dr. Jayadevi H. Chandrashekhar for her guidance in PacBio library preparation. We acknowledge Roy J. Carver Biotechnology Center for its sequencing services. We acknowledge Dr. Ashley Ceniceros, Dr. Ann Frederick, and Dr. Bruce Wellman from Carle Health for providing clinical bacteria isolates. We acknowledge Dr. Albert Cox, Dr. Essam El-Naggar, Dr. Kamlesh Patel, and Dr. Kaylyn Patterson from the Metropolitan Water Reclamation District of Greater Chicago for providing WWTP sewage samples.

Supplementary materials:

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2024.123056.

References

Abia Akebe, L.K., Sibanda, T., Selvarajan, R., El-Liethy, M.A., Kamika, I., 2022. Editorial: Environmental Reservoirs of Antibiotic Resistance Determinants: A Ticking Time Bomb for the Future Emergence of Super-Bugs of Environmental and Public Health

Alcock, B.P., et al., 2020. CARD 2020: antibiotic resistome surveillance with the Comprehensive antibiotic Resistance database. Nucleic Acids. Res. 48, D517-D525. Alcock, B.P., et al., 2023. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database.

Auerbach, E.A., Seyfried, E.E., McMahon, K.D., 2007. Tetracycline resistance genes in activated sludge wastewater treatment plants. Water. Res. 41, 1143–1151.

Bengtsson-Palme, J., et al., 2023. Towards monitoring of antimicrobial resistance in the environment: for what reasons, how to implement it, and what are the data needs?
Environ. Int. 178, 108089.

Boyle, E.A., et al., 2021. Quantification of Cas9 binding and cleavage across diverse guide sequences maps landscapes or target engagement. Sci. Adv. 7. Castanheira, M., Simner, P.J., Bradford, P.A., 2021. Extended-spectrum β-lactamases: an update on their characteristics, epidemiology and detection. JAC. Antimicrob. Resist. 3.

Chen, B., et al., 2016. Metagenomic analysis revealing antibiotic resistance genes (ARGs) and their genetic compartments in the Tibetan environment. Environ. Sci. Technol. 50, 6670–6679.
Crossette, E., et al., 2021. Metagenomic quantification of genes with internal standards. mBio 12, 1–12.
Danecek, P., et al., 2021. Twelve years of SAMtools and BCFtools. Gigascience 10.
Davis, B.C., et al., 2023. Recommendations for the use of metagenomics for routine monitoring of antibiotic resistance in wastewater and impacted aquatic environments. Crit. Rev. Environ. Sci. Technol. 53, 1731–1756.
de Abreu, V.A.C., Perdigo, J., Almeida, S., 2020. Metagenomic approaches to analyze antimicrobial resistance: an overview. Front. Genet. 11.
Debroas, D., Siguret, C, 2019. Viruses as key reservoirs of antibiotic resistance genes in the environment. ISMe J. 13, 2856.
Diamond, M.B., et al., 2022. Wastewater surveillance of pathogens can inform public health responses. Nat. Med. 28, 1992–1995, 2022 28:10.
Dong, C., et al., 2024. Dynamic migration and risk of cephalosporin antibiotic resistance genes: Move from pharmaceutical plants through wastewater treatment plants to coastal tidal flats sediments. Water. Res. 261, 121983.
Fang, H., Wang, H., Cai, L., Yu, Y., 2015. Prevalence of antibiotic resistance genes and bacterial pathogens in long-term manured greenhouse soils as revealed by metagenomic survey. Environ. Sci. Technol. 49, 1095–1104.
Fassy, J., et al., 2021. Versatile and flexible microfluidic qPCR test for high-throughput SARS-CoV-2 and cellular response detection in nasopharyngeal swab samples. PLoS. One 16, e0243333.
Ferreira, C., Otani, S., Aarestrup, F.M., Manaia, C.M., 2023. Quantitative PCR versus metagenomics for monitoring antibiotic resistance genes: balancing high sensitivity and broad coverage. FEMS. Microbes. 4.
Fuhrmeister, E.R., et al., 2024. Context-Seq: CRISPR-Cas9 Targeted Nanopore Sequencing for Transmission Dynamics of Antimicrobial Resistance. bioRxiv. https://doi.org/10.1101/2024.09.12.612745.
Garner, E., et al., 2018. Metagenomic Characterization of Antibiotic Resistance Genes in Full-Scale Reclaimed Water Distribution Systems and Corresponding Potable Systems. Environ. Sci. Technol. 52, 6113–6125.
Gaudin, M., Desnues, C., 2018. Hybrid capture-based next generation sequencing and its application to human infectious diseases. Front. Microbiol. 9, 426578.
Ghaly, T.M., Chow, L., Asher, A.J., Waldron, L.S., Gillings, M.R., 2017. Evolution of class 1 integrons: Mobilization and dispersal via food-borne bacteria. PLoS. One 12, e0179169.
Gilpatrick, T., et al., 2020. Targeted nanopore sequencing with Cas9-guided adapter ligation. Nat. Biotechnol. 38, 433–438, 2020 38:4.
Ginzinger, D.G., 2002. Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. Exp. Hematol. 30, 503–512.
Gniadkowski, M., 2008. Evolution of extended-spectrum beta-lactamases by mutation. Clin. Microbiol. Infect. 14 (Suppl 1), 11–32.
Grace, A., Sahu, R., Owen, D.R., Dennis, V.A., 2022. Pseudomonas aeruginosa reference strains PAO1 and PA14: A genomic, phenotypic, and therapeutic review. Front. Microbiol. 13, 1023523.
Guitor, A.K., et al., 2020. Capturing the resistome: a targeted capture method to reveal antibiotic resistance determinants in metagenomes. Antimicrob. Agents Chemother 64.
Gweon, H.S., et al., 2019. The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples. Environ. Microbiomes 14, 1–15.
Han, R., et al., 2020. Dissemination of Carbapenemases (KPC, NDM, OXA-48, IMP, and VIM) Among Carbapenem-Resistant Enterobacteriaceae Isolated From Adult and Children Patients in China. Front. Cell Infect. Microbiol. 10, 535376.
Harshitha, R., Arunraj, D.R., 2021. Real-time quantitative PCR: A tool for absolute and relative

quantification. Biochemistry and Molecular Biology Education 49, 800–812.
Jacky, E., et al., 2021. Robust imuchanner encoding for highly multiplexed quantitative PCR. Anal. Chem. 93, 4208–4216.
Kent, W.J., 2002. BLAT-the BLAST-like alignment tool. Genome Res. 12, 656-664.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359, 2012 9:4.
Larsson, D.G.J., Flach, C.F., 2021. Antibiotic resistance in the environment. Nature Reviews Microbiology 20, 257–269, 2021 20:5.
Lee, S.J., et al., 2023. Evaluation of two high-throughput genotyping systems for rapid identification of Canadian wheat varieties. Canadian Journal of Plant Science 103, 422–425
Li, H., et al., 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25, 2078–2079.
Li, W., et al., 2022. Population-based variations of a core resistome revealed by urban sewage metagenome surveillance, Environ. Int. 163, 107185.
Li, Y.M., et al., 2020. Metagenomic characterization of bacterial community and antibiotic resistance genes in representative ready-to-eat food in southern China. Sci. 10, 1, 14, 2020; 10: 1
Liang, X., et al., 2015. Rapid and highly efficient mammalian cell engineering via Cas9.
Liguori, K., et al., 2022. Antimicrobial resistance monitoring of water environments: a framework for standardized methods and quality control. Environ. Sci. Technol. 56, 2014; 2015.
Lin, Z.J., et al., 2024. Deciphering Multidrug-Resistant Plasmids in Disinfection Residual
Maboni, G., et al., 2022. Three Distinct Animation Platforms Differ in Detection of Antimicrobial Resistance Genes in Long-Read, Short-Read, and Hybrid Sequences. Derived from Total Genomic DNA or from Purified Plasmid DNA. Antibiotics 11, 1400.
Madden, T.L. Chapter 16: The BLAST Sequence Analysis Tool. (2013).
Mao, Y., Nguyen, T.H., 2023. Multiplexed CRISPR-based target-enriched next-generation sequencing for detecting antibiotic resistance genes in environmental samples v2.
https://doi.org/10.17504/protocolso.io.8epv5xdnjg1b/v2.
Y. Mao et al.
Water Research 274 (2025) 123056
Mao, X., et al., 2024. Standardization in global environmental antibiotic resistance genes (ARGs) surveillance. Crit. Rev. Environ. Sci. Technol. https://doi.org/10.1080/10643389.2024.2344453.
McCarty, N.S., Graham, A.E., Studená, L., Ledesma-Amaro, R., 2020. Multiplexed CRISPR technologies for gene editing and transcriptional regulation. Nat. Commun. 11, 1–13, 2020 11:1.
McCombie, W.R., McPherson, J.D., 2019. Future promises and concerns of ubiquitous next-generation sequencing. Cold. Spring. Harb. Perspect. Med. 9.
McKenna, M., 2013. Antibiotic resistance: The last resort. Nature 499, 394–396.
Olson, R.D., et al., 2023. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. Nucleic. Acids. Res. 51, D678–D689.
Olwagen, C.P., Adrian, P.V., Madhi, S.A., 2019. Performance of the Biomark HD real-time qPCR System (Fluidigm) for the detection of nasopharyngeal bacterial pathogens and Streptococcus pneumoniae typing. Sci. Rep. 9, 1–11, 2019 9:1.
Pagani, L., et al., 2004. Multifocal Detection of Multidrug-Resistant Pseudomonas aeruginosa Producing the PER-1 Extended-Spectrum β-Lactamase in Northern Italy. J. Clin. Microbiol. 42, 2523.
Poirel, L., et al., 2001. GES-2, a class A beta-lactamase from Pseudomonas aeruginosa with increased hydrolysis of imipenem. Antimicrob. Agents Chemother 45, 2598–2603.
Pricing | Biotech. https://biotech.illinois.edu/htdma/pricing.
Quan, J., et al., 2019. FLASH: a next-generation CRISPR diagnostic for multiplexed detection of antimicrobial resistance sequences. Nucleic. Acids. Res. 47, E83.
Ruby, J.G., Bellare, P., DeRisi, J.L., 2013. PRICE: Software for the targeted assembly of components of (Meta) genomic sequence data. G3: Genes, Genomes, Genetics 3, 865–880.

Samorodnitsky, E., et al., 2015. Evaluation of hybridization capture versus amplicon-based methods for whole-exome sequencing. Hum. Mutat. 36, 903–914.
Smith, T., Heger, A., Sudbery, I., 2017. UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. Genome Res. 27, 491–499.
Sratnongheam, 1., et al., 2024. High throughput qPCR unveis shared antibiotic resistance genes in tropical wastewater and river water. Sci. Total Environ. 908, 167867.
Sundararaman, B., et al., 2023. A hybridization target enrichment approach for pathogen genomics. mBio 14.
United Nations Environment Programme. Bracing for superbugs: strengthening environmental action in the one health response to antimicrobial resistance. (2023).
Wang, C., et al., 2024b. Metagenomic absolute quantification of antibiotic resistance genes and virulence factor genes-carrying bacterial genomes in anaerobic digesters. Water. Res. 253, 121258.
Wang, Z., Cai, M., Du, P., Li, X., 2024a. Wastewater surveillance for antibiotics and resistance genes in a river catchment: Spatiotemporal variations and the main drivers. Water. Res. 251, 121090.
Xie, N.G., et al., 2022. Designing highly multiplex PCR primer sets with simulated annealing design using dimer likelihood estimation (SADDLE). Nat. Commun. 13, 1–10, 2022 13:1.
Yin, X., et al., 2019. Exploration of the antibiotic resistome in a wastewater treatment plant by a nine-year longitudinal metagenomic study. Environ. Int. 133, 105270.
Zaheer, R., et al., 2018. Impact of sequencing depth on the characterization of the microbiome and resistome. Sci. Rep. 8, 1–11, 2018 8:1.
Zhang, Q., Ishii, S., 2018. Improved simultaneous quantification of multiple waterborne pathogens and fecal indicator bacteria with the use of a sample process control. Water. Res. 137, 193–200.
Zhou, Z., et al., 2021. Metagenomic analysis of microbiota and antibiotic resistome in household activated carbon drinking water purifiers. Environ. Int. 148, 106394.