Environmental Science & Technology

pubs.acs.org/est

Article

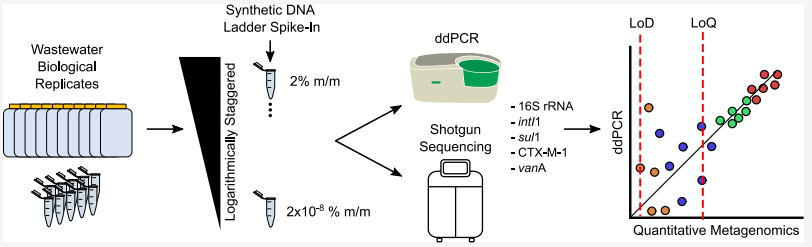
Evaluating Quantitative Metagenomics for Environmental Monitoring of Antibiotic Resistance and Establishing Detection Limits

Benjamin C. Davis,* Peter J. Vikesland, and Amy Pruden

Cite This: *Environ. Sci. Technol.* 2025, 59, 6192−6202

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Metagenomics holds promise as a comprehensive, nontargeted tool for environmental monitoring. However, one key limitation is that the quantitative capacity of metagenomics is not well-defined. Here, we demonstrated a quantitative metagenomic technique and benchmarked the approach for wastewater-based surveillance of antibiotic resistance genes. To assess the variability of low-abundance oligonucleotide detection across sample matrices, we spiked DNA reference standards (meta sequins) into replicate wastewater DNA extracts at logarithmically decreasing mass-to-mass percentages (m/m%). Meta sequin ladders exhibited strong linearity at input concentrations as low as 2×10^{-3} m/m% ($R^2 > 0.95$), with little to no reference length or GC bias. At a mean sequencing depth of 94 Gb, the limits of quantification (LoQ) and detection were calculated to be 1.3×10^3 and 1 gene copy per μL DNA extract, respectively. In wastewater influent, activated sludge, and secondary effluent samples, 27.3, 47.7, and 44.3% of detected genes were ≤LoQ, respectively. Volumetric gene concentrations and log removal values were statistically equivalent between quantitative metagenomics and ddPCR for 16S rRNA, *intI*1, *sul*1, CTX-M-1, and *van*A. The quantitative metagenomics benchmark here is a key step toward establishing metagenomics for high-throughput, nontargeted, and quantitative environmental monitoring.

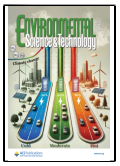
KEYWORDS: *environmental monitoring, antibiotic resistance, quantitative metagenomics, internal standards, limit of quantification, limit of detection*

INTRODUCTION

The need for environmental monitoring approaches that broadly capture public health threats was exemplified by the COVID-19 pandemic. In response to the pandemic, wastewater-based surveillance (WBS) infrastructure expanded exponentially across the globe, presenting the opportunity to expand such assets to capture a multitude of agents of concern.[1−3] High-throughput, quantitative, nontargeted methods that can capture multiple targets would present a transformative advancement to the field of environmental monitoring. Such an approach would be particularly useful for monitoring antimicrobial resistance (AMR), e.g., wherein thousands of known, functionally verified antibiotic resistance genes (ARGs) in public databases[4] can be simultaneously screened. Droplet digital polymerase chain reaction (ddPCR) and quantitative PCR (qPCR) were quickly established as the gold standard for monitoring of SARS-CoV-2 and polio virus[5,6] and have also been widely applied for sensitive quantification of specific ARGs in surface water,[7,8] recycled water,[9] and wastewater systems.[10] ddPCR is proving to be a superior means of quantifying specific genes of interest in a given sample, with high sensitivity, reduced vulnerability to inhibitors, and a broad dynamic range of quantification.[11] Realistically, however, ddPCR is only capable of detecting a handful of genes at a time.

Recent advances in microfluidic-based high-throughput qPCR (HT-qPCR) have enabled the simultaneous detection of hundreds of genes and have seen particular application to ARG monitoring. However, such approaches widely lack quality control and validation metrics (e.g., standard and melt curves). This results in uncertainty in individual target detection and quantification.[12] Additionally, a priori identification of gene targets requires the design and validation of

Received: August 12, 2024
Revised: January 13, 2025
Accepted: January 14, 2025
Published: March 18, 2025



https://doi.org/10.1021/acs.est.4c08284
Environ. Sci. Technol. 2025, 59, 6192−6202

suitable primers, significantly hindering the capacity to detect emerging gene variants in real-time. Shotgun metagenomic sequencing represents a distinct advantage in this regard, as it provides high-throughput nontargeted profiling of the multitude of genomes comprising a microbial community.[13] Recent studies have demonstrated metagenomics as a promising nontargeted means for ARG monitoring in wastewater[14−16] and there is currently momentum in the application of metagenomics for comprehensive WBS of ARGs at global scales.[17,18] However, the quantitative capacity of metagenomic data has remained in question due to its inherently compositional nature.[19] This has limited the overall comparability of metagenomic data across studies as well as its value for environmental and epidemiological applications, including statistical analysis and risk assessment.[20,21]

Evaluating metagenomic data in terms of relative abundances, i.e., normalization to a common denominator, such as a housekeeping gene or the total reads yielded from a sequencing run, is a common means of facilitating comparison of metagenomes across studies.[22,23] However, absolute concentrations that normalize metagenomic gene counts per sample volume basis would be of broader value, e.g., for calculation of removal rates and exposure doses. Due to the random sequencing of all genomic DNA across complex microbiomes, there is uncertainty with regards to the detection limits for various genes of interest. Statistical thresholds are needed to define the limits of quantification (LoQ) and detection (LoD) in environmental matrices. Through systematic analysis of publicly available literature and associated data, we recently estimated that a sequencing depth approaching 100 giga base pairs (Gb) would be necessary to consistently achieve a theoretical metagenomic coverage ≥0.90 in most wastewater matrices.[22] We expect that a sequencing depth ∼100 Gb per wastewater DNA extract should maximize the probability of detecting extremely low abundance oligonucleotides in these compositional data sets.

The goal of this study was to demonstrate and evaluate a quantitative metagenomics technique for deriving volumetric concentrations of genes of interest from metagenomic data and explore the limitations of second-generation short-read sequencing for detecting low-abundance genes. As proposed by Hardwick et al. 2018,[24] synthetic nucleic acid reference standards, i.e., "sequins," were spiked into replicate DNA extracts as internal standards. Sequins are designed to represent a range of features and complexity encountered in natural microbial communities, while sharing no homology to RefSeq sequences.[25] The sequins employed in this study, meta sequins, were designed specifically for metagenomic application and consist of a mixture of 86 unique DNA oligonucleotides of varying lengths (987−9120 bp) and GC content (24−71%) that are present at 16 discrete input proportions, thus forming a ladder.[25] Meta sequins were originally benchmarked for intersample normalization as well as for measuring fold changes between microbial communities. Previous studies have attempted to advance quantitative metagenomics through spiking exogenous whole genomes for ARG quantification,[26] but genetic homology between naturally derived spike-ins and environmental DNA cross-detection adds uncertainty to quantification. To overcome this bias, synthetic DNA oligonucleotides with embedded xenobiotic insertions (consecutive stop codons) were recently benchmarked,[27] but their small insertion sizes (103−430 bp) significantly reduce the probability of detecting these references at extremely low abundance (e.g., <100 gc/μL).

The specific objectives of this study were to (1) evaluate the behavior of meta sequin ladders at decreasing spike-in mass-to-mass percentages (m/m%), (2) establish the technical limitations of Illumina sequencing for detecting low abundance oligonucleotides by defining the LoQ and LoD, and (3) compare the efficacy of sequin standards for quantifying ARGs in different wastewater sample types through direct comparison to ddPCR. Meta sequins were spiked at logarithmically decreasing m/m% into replicate DNA extracts of influent, activated sludge, and secondary effluent. These reflect three wastewater matrices that represent distinct levels of microbial community complexity and expected ARG composition and concentration. Our overall approach establishes the quantitative capacity of metagenomic sequencing for environmental monitoring of genes of interest, including consideration of sample matrix, internal standard concentration, and sequencing depth. The resulting protocol is expected to be of value to ongoing efforts to establish shotgun metagenomics as a quantitative monitoring tool.

MATERIALS AND METHODS

Sample Collection, DNA Extraction, and Purification. Grab samples were collected in November 2021 from a local 5 MGD conventional WWTP in Blacksburg, Virginia. Ten Influent, 10 activated sludge, and 10 secondary effluent samples were collected in individual 50 mL, 50 mL, and 500 mL autoclaved polypropylene bottles, respectively and transported on ice to the laboratory for processing within 2 h. Briefly, samples were vacuum filtered onto 0.45-μm mixed cellulose-ester filters, with 50, 10, and 500 mL filtered for each sample type, respectively, to create biological replicates. One hundred milliliters of deionized water were vacuumed through an additional filter to serve as the filter blank and negative control. Filters were placed in 2 mL centrifuge tubes, fixed with 1 mL 100% ethanol, and stored at −20 °C before DNA extraction. Ethanol fixed filters were fragmented with flame-sterilized tweezers and placed into lysing matrix E tubes of the FastDNA Spin Kit for Soil (MPBio, Solon, OH). Samples were then homogenized via bead-beating (40 s at 6 m/s) with the FastPrep-24 5G (MPBio), further extracted according to manufacturer's instructions, and eluted in 100 μL of elution buffer. DNA extracts were first quantified using a dsDNA high sensitivity assay kit on a Qubit Fluorometer (Invitrogen, Carlsbad, CA), and 260/280 UV−vis ratios were checked on a NanoPhotometer Pearl (Implen). Each DNA extract was purified using a ZymoBIOMICS DNA Clean & Concentrator kit, eluted with 50 μL elution buffer, and requantified and quality checked.

ddPCR of Gene Targets. All ddPCR analysis were conducted using a BioRad QX200Droplet Digital PCR system (Bio-Rad, Hercules, CA, USA). ddPCR was performed on all cleaned DNA extracts to quantify the abundance of total bacteria (16S rRNA genes),[28] *sul*1,[29] *int*I1,[30] CTX-M-1,[31] and *van*A.[32] Each assay was optimized for ddPCR by performing a thermal gradient on gBlock (IDT) standards. Optimized thermocycle conditions used are presented in Table S1. Initially, subsampled DNA from the first replicates of each wastewater matrix were aliquoted to identify dilution factors (1:100, 1:250, 1:500, 1:1000, and 1:1500) for which target concentrations were within the upper and lower detection limits and PCR inhibition was minimized. The 1:1500 and

https://doi.org/10.1021/acs.est.4c08284
Environ. Sci. Technol. 2025, 59, 6192−6202

1:100 dilutions were found to maximize output concentrations of 16S rRNA and target ARGs for all three wastewater matrices, respectively. The minimum number of accepted droplets was set following the manufacturer's recommendation (>10,000 per well) and positive droplets were thresholded one standard deviation above the negative droplet cloud. Using diluted aliquots, each ddPCR target was quantified in analytical triplicate with deionized water as the no template control (NTC).

Internal Reference Standard Spiking and Sequencing. Sequins (https://sequins.bio/) were received from the Garvan Institute of Medical Research (Sydney 2010 NSW, Australia) as lyophilized nucleic acids. Meta sequin "Mixture A" was resuspended according to the manufacturer's instructions to a concentration of 2 ng/μL using molecular grade water and quantified using a Qubit Fluorometer (Thermo Fisher, USA). Mixture A contains 86 individual sequins at 16 discrete input proportions with at least five unique sequins at each proportion level, forming a reference ladder. The details for the meta sequins can be found in Table S2. To generate a logarithmically decreasing spike-in gradient, sample DNA extracts were first normalized to ~1000 ng in 500 μL microcentrifuge tubes (Table S3). Meta sequins were serially diluted nine times, starting at 2 ng/μL, and 10 μL of each dilution were added to corresponding replicate samples. This achieved ratios of 2, 0.2, 0.02%, etc. m/m (meta sequin mass to sample mass) of sequins per DNA extract for sequencing. Manufacturer instructions recommend adding sequins at a 2% m/m ratio and was defined here as a dilution factor of 1. Library preparation (KAPA HyperPrep PCR-free workflow targeting 500 bp insert sizes) and sequencing was carried out at the Duke Center for Genomic and Computational Biology. Libraries were sequenced across all four lanes on a single NovaSeq 6000 S4 flow cell with 150 bp chemistry targeting 100 Gb per sample. The undiluted filter blank extract was below the LoQ for 16S rRNA in all analytical triplicates and was not submitted for sequencing. Metagenomic libraries have been deposited into the Sequence Read Archive under accession PRJNA1095031.

Bioinformatic Analysis. R1 and R2 files from each of the four flow cell lanes were concatenated. Adapters were removed and reads were quality filtered, trimmed, and merged using fastp with default parameters.[33] Metagenomic coverage was estimated using Nonpareil3.[34] To align reads to the sequin reference standards, the Anaquin software, a dedicated package designed to analyze sequin reference standards, was run using the "meta" option.[35] Calibration settings for each Anaquin command were kept at 0.01 as calibration files were not used in downstream analysis. The resulting count tables for meta sequin Mixture A were used to generate standard curves. NTC sequins were screened with the Anaquin software to estimate the rate of erroneous read mapping. For each meta sequin ladder, the mean reads per kilobase (RPK; eq 1) sequin count was taken per input proportion, resulting in 16-point reference curves.

$$\text{Reads per kilobase (RPK)} = \frac{\text{\#Reads aligned to reference}}{\text{Length of reference (nt)}/1000} \quad (1)$$

The meta sequin features (e.g., sequin length, GC content, input proportions), found in the source files of the Anaquin software, were used to calculate the input concentrations of each sequin (Tables S2 and S4). Merged clean reads devoid of sequin reads were then queried against the Comprehensive Antibiotic Resistance Database (CARD, v. 3.0.3, protein homologue model) using DIAMOND BLASTx (max-target-seqs = 1, query coverage = 80%, id = 80%).[4] 16S rRNA counts were determined by mapping reads to the Greengenes database[36] using minimap2[37] and the resulting bam files were summarized with SAMtools.[38] To achieve a sliding-scale for gene quantitation specificity, the resulting annotation files were subset to id thresholds of 90 and 99%. To test the effect of sequencing depth on meta sequin ladder fidelity and the LoQs and LoDs, each sample was proportionally sub set to 1, 0.75, 0.5, 0.25, 0.1, 0.05, 0.01, 0.001× the original sample sizes using SeqKit[39] and reanalyzed. All read counts were converted to RPK to account for reference sequence length biases.

Quantification of Reference Genes. To quantify ARGs in each wastewater DNA extract, log−log models were first fit to meta sequin reference ladders using eq 2:

$$\log_{10}(\text{RPK}_{\text{sequin}-i}) = \beta_1(\log_{10}(C_{\text{sequin}-i})) + \beta_0 \quad (2)$$

where ($\text{RPK}_{\text{sequin-}i}$) is the RPK normalized read count of sequin-i, β_1 is the slope, ($C_{\text{sequin-}i}$, gc/μL) is the gene copy concentration of sequin-i spiked into the DNA extract, and β_0 is the intercept. Solving eq 2, we can directly quantify target genes in each wastewater DNA extract by applying eq 3:

$$C_{\text{gene}-i}\left(\frac{\text{gc}}{\mu\text{L}}\right) = \left(10^{\log_{10}(\text{RPK}_{\text{gene}-i})-\beta_0/\beta_1}\right) \quad (3)$$

where ($C_{\text{gene-}i}$, gc/μL) is the concentration of gene-i per μL of DNA extract, ($\text{RPK}_{\text{gene-}i}$) is the RPK normalized read count of gene-i, and the β_0 and β_1 parameters are derived from the fitted models in eq 2. To convert ($C_{\text{gene-}i}$, gc/μL) into a per-volume-of-sample basis, the total DNA extraction volume (EV_{total}, μL), the DNA extraction used (EV_{used}, μL) (i.e., submitted for library prep and sequencing), and the sample volume filtered (FV, mL) were applied as in eq 4:

$$C_{\text{gene}-i}\left(\frac{\text{gc}}{\text{mL}}\right) = \frac{C_{\text{gene}-i}\left(\frac{\text{gc}}{\mu\text{L}}\right) x \text{EV}_{\text{total (uL)}}}{\text{EV}_{\text{used (uL)}} x \text{FV (mL)}} \quad (4)$$

The sequencing yield for each individual sequin ($Y_{\text{seq-}i}$; eq 4) was used to evaluate recovery of entire meta sequin ladders at each dilution factor. $Y_{\text{seq-}i}$ is a unitless parameter that relates the mass ratio of spiked-in oligonucleotides $\left(\frac{M_{\text{seequin}-i}}{M_{\text{Total}}}\right)$ to the total recovered base pairs $\left(\frac{bp_{\text{sequin}-i}}{bp_{\text{Total}}}\right)$ as described by[27]

$$\left(\frac{M_{\text{seequin}-i}}{M_{\text{Total}}}\right) x Y_{\text{seq}-i} = \left(\frac{bp_{\text{sequin}-i}}{bp_{\text{Total}}}\right) \quad (5)$$

where the total sequencing yield (Y_{seq}) is the sum of all $Y_{\text{seq-}i}$ values for each meta sequin ladder in each sample.

Statistical Analyses. Before any statistical analyses were performed, ddPCR and quantitative metagenomic data were ($\log_{10}$ + 1) transformed and tested for normality using the Shapiro-Wilk test. Data were found to be non-normal, warranting nonparametric testing. A Kruskal−Wallis test was used to assess differences in ladder viability as a function of dilution factor and sequencing depth. Quantitative metagenomic and ddPCR gc/mL values were tested for equivalence using Wilcoxon Two One-Sided Tests (TOST) with stand-

https://doi.org/10.1021/acs.est.4c08284
Environ. Sci. Technol. 2025, 59, 6192−6202

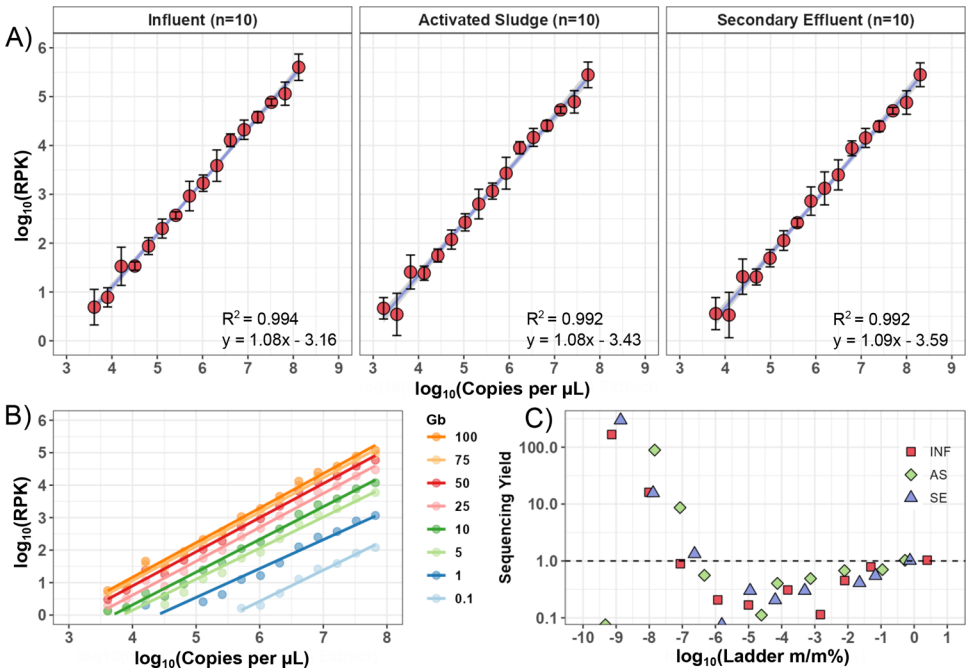


Figure 1. Performance of sequin ladders across sample matrices, sequencing depths, and spike-in percentages. (A) Linearity of meta sequin "Mixture A" spiked at 2% m/m, the manufacturer recommended spike-in concentration, in different wastewater matrices. Each point represents the mean log-transformed RPK sequin count per input proportion (see Table S1), resulting in 16-point reference curves. Error bars represent ± standard deviations of log-transformed data. (B) Ladder linearity at 2% m/m in an influent sample subsampled to discrete sequencing depths in Gb. Error bars are not shown for simplicity. Log−log models for each subset can be found in Table S5. (C) Total sequencing yield as a function of decreasing input m/m% and sample matrix.

ardized effect sizes (Cohen's d) for each target in each matrix used to set equivalence bounds.[40] All statistical analyses including linear regression and figure generation were performed in R 4.41[41] using the packages tidyverse, ggplot2, and pheatmap. Figures were paneled and annotated in Inkscape.

RESULTS

Meta Sequin Detection Patterns across Spike-In Concentrations. After read quality filtering and trimming, an average of 6.3×10^8 reads (94 Gb) were generated per sample, resulting in a total library size of 2.8 Tb. Total per sample reads derived from meta sequins ranged from a single read to 7.5×10^6 reads (Table S3). Metagenomic coverage approximations were consistent across replicates and approached saturation for influent (mean ± standard deviation; 0.807 ± 0.016), activated sludge (0.862 ± 0.001), and secondary effluent samples (0.850 ± 0.013) (Table S3). Strong linearity ($R^2 > 0.99$) between input sequin copies/μL DNA extract and mean RPK values were observed across ladder m/m% and sample matrices (Figures 1A and S1). All 86 sequins were detected at the first two dilutions across the three sample matrices, and at least a single sequin was detected in 27/30 samples (Figure S2). At 2% m/m, the recommended manufacturer spike-in concentration, input copies ranged from 5.1×10^3 copies/μL to 1.65×10^8 copies/μL, a 4.5-log range. Log−log models across the three sample matrices at 2% m/m were nearly identical and displayed an R^2 of 0.993 ± 0.001 and slope of 1.08 ± 0.006 (Figure 1A and Table S4). Other metagenomic studies utilizing exogenous DNA standards have reported similarly strong linear relationships between predicted and expected concentrations of spiked oligonucleotides.[25−27,42,43] Further, to test the effect of sequencing depth on the ladder linearity, we subset the influent samples to eight discrete depths and found consistent R^2 values and slopes at 2% m/m and depths at or above 5 Gb (Figure 1B and Table S5). At spike-in percentages $\leq 2 \times 10^{-3}$% m/m, ladder recovery became unreliable at sequencing depths <25 Gb and can be considered a lower limit. No viable ladders could be recovered below 2×10^{-4}% m/m at any sequencing depth (Figure S3 and Table S4).

For ladder spike-in percentage $\geq 2 \times 10^{-4}$% m/m, we found the 16-point reference curve RPK values between influent, activated sludge, and secondary effluent samples to be statistically indistinguishable (Kruskal−Wallis, $p > 0.05$), indicating that changes in the inherent nucleic acid complexity of the DNA extracts did not influence detection of the reference sequences (Figure S4). We also observed little to no reference length (R^2 = 0.00154, p = 0.2434) or GC (R^2 = 0.00039, p = 0.2558) bias in the sequencing yield of individual oligonucleotides across the entire experiment (Figure S5). Sequencing recovery of the ladders, measured as total detected base pairs over total input base pairs, was also remarkably stable across sequencing depths and sample matrices (Figure S6). Therefore, these observations allowed us to treat influent, activated sludge, and secondary effluent meta sequin ladders at each dilution factor as technical replicates, thus increasing statistical power in downstream analyses.

Y_{seq} values, measured as the mass ratio of spiked-in oligonucleotides to the total recovered base pairs (ideal value of 1), reached saturation at the highest input m/m% and were stable across matrices (1.02 ± 0.009). This stability was not maintained at subsequent dilution factors, and values steadily declined to a minimum at 2×10^{-4} m/m% (0.193 ± 0.097) as the detection of individual sequins became sporadic. A recovery in Y_{seq} was then observed as read counts began to

https://doi.org/10.1021/acs.est.4c08284
Environ. Sci. Technol. 2025, 59, 6192−6202

surpass calculated input sequin concentrations due to random detection of sequins at extremely low concentrations (Figure 1C).

Defining and Establishing the LoQ and LoD. LoQ and LoD were established by analyzing the coefficients of variation (CV%) of Y_{seq} for individual sequins as a function of calculated input copies, with samples across the three wastewater matrices at each m/m% treated as technical replicates (Figure 2A).

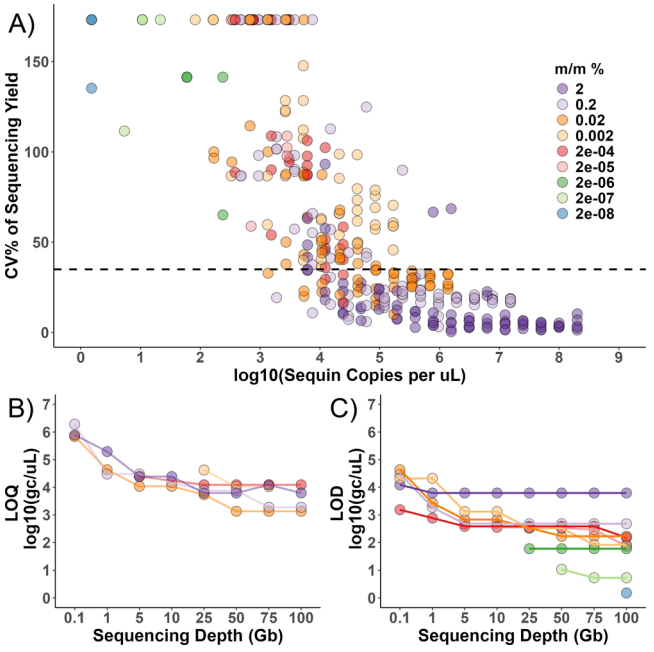


Figure 2. The LoQ and LoD of quantitative metagenomics. (A) Coefficient of variation (CV%) of the sequencing yield of all 86 sequins across ladder spike-in m/m%. Influent, activated sludge, and secondary effluent samples at each m/m% were treated as technical replicates. The black dashed line marks the generally recommended threshold CV = 35% for determining the LoQ of qPCR experiments. (B) The calculated LoQ and (C) LoD of quantitative metagenomics as a function of sequencing depth and ladder spike-in m/m%.

Following general guidelines for qPCR,[44] we stringently defined the LoQ as the lowest input sequin concentration that was detected across all three technical replicates with a Y_{seq} CV $\leq$ 35%, and the LoD as the lowest individual sequin concentration detected across all three technical replicates. Based on these criteria, the overall LoQ of the experiment at ~100 Gb was found to be 1350 gene copies (gc)/ μL in the 0.02% m/m replicates (Figure 2A). The LoD was found to be 1 gc/μL, which occurred in the 2 $\times$ 10^{-8}% m/m replicates. There was never an instance of a sequin being detected at a calculated input copy number <1 gc/μL, thus indicating strong agreement between derived concentrations and the physical spike-in of individual sequins. After screening for internal sequin NTCs for Mix A, a singular reference (ID: S1067_MR_008_A, input proportion = NA) was found to be erroneously aligned to across all sample matrices (0−981 reads/sample) (Figure S7).

We further explored the LoQ and LoD as a function of sequencing depth on each of the spike-in ladder concentrations. We observed steadily decreasing LoQs and LoDs with increasing sequencing depth (Figure 2B,C). Only for samples with ladder spike-in percentages $\geq 2 \times 10^{-5}$% could an LoQ be established due to sporadic detection of oligonucleotides at concentrations less than ~10^4 gc/μL across replicates (Figure 2A and Table S4). For the three highest dilution factors, the LoQ was proportionally reduced an order of magnitude between the 10 Gb subsample and at 100 Gb (mean 1.7 $\times$ 10^4 versus 3.1 $\times$ 10^3 gc/μL) (Figure 2B). At ~100 Gb, the LoQ of the 2, 0.2, and 0.02% m/m concentrations were 6203, 1875, and 1350 gc/μL, respectively. The corresponding LoDs for these m/m concentrations were 6203, 476, and 169 gc/μL. The LoDs at each spike-in percentage were approximately an order of magnitude less than their corresponding LoQs and were largely thresholded by the minimum oligonucleotide concentration for a given dilution factor (Figure 2C and Table S4).

High Throughput Quantification of ARGs via Quantitative Metagenomics. The absolute gene copies of ARGs across all 30 samples were estimated using the mean slope

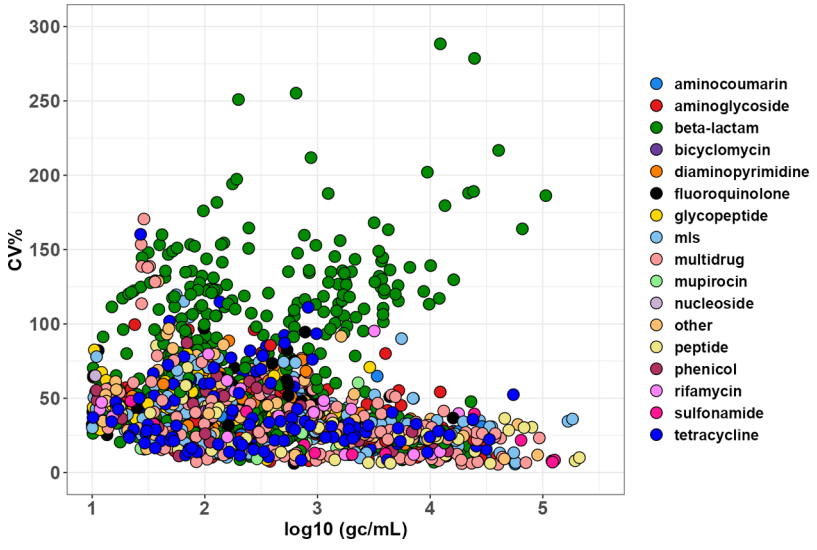


Figure 3. Variable detection of unique ARGs across concentration range. ARGs across all 30 samples were quantified using the mean slope (1.08) and intercept (−3.39) of the log−log models with 2% m/m ladders at ~100 Gb (CARD v3.0.3, ID = 80%, coverage = 80%). Each point represents a unique ARG detected across matrix replicates (n = 10) color coded by antibiotic class. Singletons are omitted.

https://doi.org/10.1021/acs.est.4c08284
Environ. Sci. Technol. 2025, 59, 6192−6202

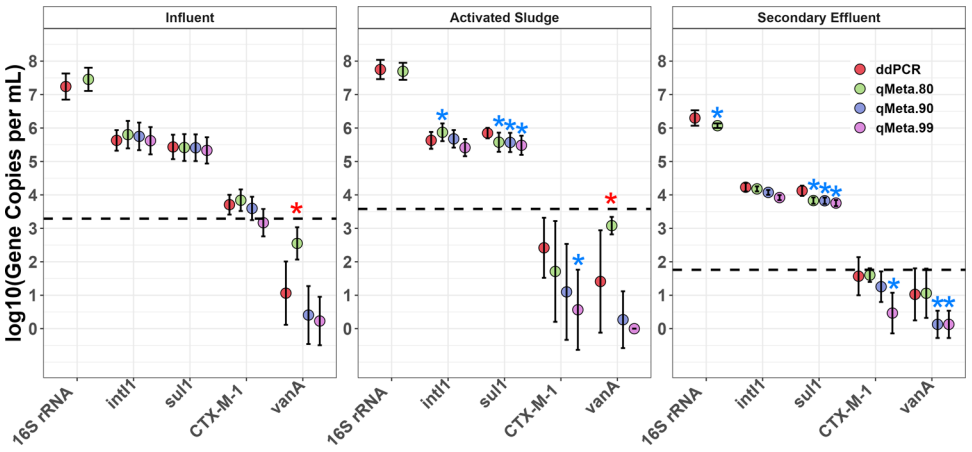


Figure 4. Comparison of ddPCR and quantitative metagenomic ARG quantification. Dot plot of quantitative metagenomic gene concentrations were determined at 80, 90, and 99% amino acid identity (i.e., qMeta.80, qMeta.90, qMeta.99) and 80% query coverage. All ten biological replicates were quantified for ddPCR and quantitative metagenomics for each matrix. Error bars represent ± standard deviations. Blue * = compared to ddPCR, the data are statistically equivalent, but the effect size is greater than zero. Red * = compared to ddPCR, the data are not statistically equivalent, and the effect size is greater than zero. See Table S6 for wilcox_TOST test results. The horizontal dashed lines represent the LoQs for each sample matrix.

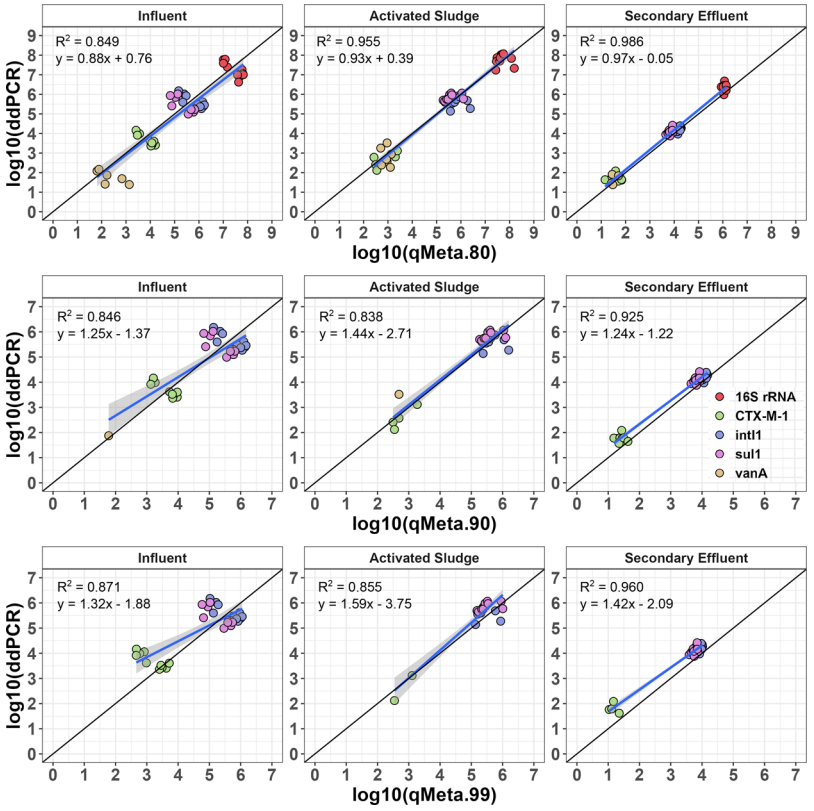


Figure 5. Correlations between ddPCR and quantitative metagenomics ARG concentrations. Quantitative metagenomic gene concentrations were determined at 80, 90, and 99% amino acid identity (i.e., qMeta.80, qMeta.90, qMeta.99) and 80% query coverage. Solid black diagonal line represents ideal slope of 1. Blue line represents results of linear regression with 95% confidence intervals.

(1.08) and intercept (−3.39) of the 2% m/m log−log models (Figure 1A). For example, applying these parameters to eq 3, 1 RPK (e.g., a single read aligning to a 1 kb reference) equates to 1377 gc/uL in each extract. We then aligned all reads to the CARD database and detected 901 ± 56, 493 ± 37, and 634 ± 32 unique ARGs across the influent, activated sludge, and secondary effluent samples, resulting in total measured $\log_{10}$ concentrations of 7.5 ± 0.4, 7.1 ± 0.3, and 5.5 ± 0.2 gc/mL, respectively. The most abundant ARG classes detected were beta-lactams, macrolide-lincosamide-streptogramin (MLS), multidrug, peptide, aminoglycosides, and tetracyclines (Figure S8). Parallel to the sequin standards, when quantifying individual ARGs between replicates, we observed a steady increase in the CV% as gene concentrations decreased (Figure 3). Beta-lactam ARGs, specifically variants of the TEM subtype, displayed the highest CV's (mean = 111.1%), followed by bicyclomycin (*bcr*-1, 53.6%), fluoroquinolone (QnrB/D variants, 40.9%), and tetracyclines (36.9%). The

https://doi.org/10.1021/acs.est.4c08284
Environ. Sci. Technol. 2025, 59, 6192−6202

exacerbated variation in calculated gene concentrations for the beta-lactamases, even between 10 biological replicates, highlights the inherent limitations of "best-hit" gene counting strategies where a read can be mapped to multiple closely related variants, but only one can be chosen and tallied.

After accounting for individual DNA extraction and filtration volumes, the mean LoQ of the 2% m/m standards (6203 gc/μL) spiked into influent, activated sludge, and secondary effluent samples were found to be 1986, 3837, and 56 gc/mL, respectively. We then applied these thresholds to each of the ten replicates per matrix to determine the proportion of unique ARGs that were <LoQ (Figure S9). For the detection of individual ARGs, we found that on average 27.3, 47.7, and 44.3% of detected genes were ≤LoQ for influent, activated sludge, and secondary effluent samples, respectively at ∼100 Gb. The complete list of detected ARGs and their estimated concentrations can be found in Table S7.

Comparison of Quantitative Metagenomics and ddPCR. The quantitative metagenomic approach was benchmarked by comparing gene concentration estimates (gc/mL) to ddPCR measurements of five target genes: 16S rRNA, *int*I1, *sul*1, CTX-M-1, and *van*A. Based on previous observations of the studied treatment plant,[15] these genes represent a typical range of concentrations in wastewater, some of which approach the LoQs and LoDs of each method. For quantitative metagenomic gene calculations, all ten biological replicates per matrix were analyzed using the average slope and intercept of the 2% m/m log−log models and metagenomic reads mapped to each gene's corresponding reference in the CARD database. We found that 37/39 (94.8%) comparisons between quantitative metagenomics and ddPCR $\log_{10}$ transformed gc/mL concentrations were statistically equivalent (Wilcoxon TOST; TOST $p < 0.05$) in value, even at varying degrees of alignment specificity (Figure 4 and Table S6). However, for 12/39 (30.8%) comparisons the gene concentrations were statistically equivalent, but the effect size (i.e., measured difference) between them was statistically greater than zero (TOST $p < 0.05$; NHST $p < 0.05$; blue * in Figure 4). In other words, the data provided strong evidence that the effect size was not zero, yet this nonzero effect was still small enough to be considered negligible. These statistical equivalencies with negligible differences were largely observed for *sul*1 measurements in activated sludge ($\log_{10}$ ddPCR: 5.85 ± 0.16 vs $\log_{10}$ qMeta.80:5.78 ± 0.29) and secondary effluent samples (4.12 ± 0.15 vs 3.83 ± 0.095), as well as CTX-M-1 and *van*A concentrations at or below the LoQ. For two quantitative metagenomic *van*A measurements at 80% amino acid identity, the measured values were both not equivalent to ddPCR and the effect size was greater than zero (red * in Figure 4). It is worth noting that *int*I1, *sul*1, and CTX-M-1 influent concentrations observed in this study are directly comparable to those derived by qPCR from WWTPs worldwide.[45]

We further compared ddPCR to derived quantitative metagenomics concentrations using simple linear regression across all targets. We found strong correlations ($R^2 > 0.85$) between measurements that varied across wastewater matrix and alignment specificity (Figure 5). Overall, the secondary effluent samples displayed the greatest linearity, presumably due to the lower biological variability of this matrix. Generally, linearity strength decreased as specificity increased due to the loss of detection of low abundance ARGs. Despite differences in quantitation of individual gene occurrences, log reduction values (influent vs secondary effluent) estimated via ddPCR versus quantitative metagenomics at 80% ID (qMeta.80) measurements were statistically equivalent, with negligible effect sizes across all targets (Figure S10). Overall, the coefficient of variation of measured gene concentrations across all assays by ddPCR versus qMeta.80 were statistically equivalent (2.14 ± 0.14 vs 2.17 ± 0.02; $p = 0.00023$).

DISCUSSION

This study demonstrated quantitative metagenomics as a viable approach for high-throughput, nontargeted quantitation of genes of interest in environmental samples, examining ARGs in three different wastewater matrices (influent, activated sludge, and effluent) as an exemplar. Addition of internal spike-in DNA standards provided a means to calculate ARG concentrations in units of gc/mL, which is much more amenable to informing epidemiological and risk assessment models as well as calculating removal rates achieved by unit operations. Benchmarking to ddPCR provided comparison to the most sensitive method currently available for quantifying individual gene targets.[46,47] While small discrepancies in quantification of specific genes (e.g., *sul*1) were noted between quantitative metagenomics and ddPCR, the overall technical variation between the methods were statistically equivalent. Further, the estimated log removal values derived from both methods were comparable. Given that this study was performed using multiple biological replicates of wastewater samples, it is to be expected that the inherent diversity of the samples themselves contribute to the observed variability in measurements.

In general, calculated gene concentrations were most comparable for targets ≥LoQ for quantitative metagenomics, approximately 10^3 gc/μL for this experiment at ∼100 Gb. These thresholds were also dependent upon ladder spike-in concentrations. Similar quantitative thresholds have been documented previously; for example, Li et al.[27] observed an LoQ of their quantitative metagenomics approach of approximately $10^{3.5}$ gc/μL at a sequencing depth of ∼6 Gb, and Langenfeld et al.[43] calculated the LoD for their own meta sequin ladders to be approximately 500 gc/μL at ∼30 Gb. These results indicate the possibility of diminishing returns regarding efforts to decrease the LoQ/LoD of quantitative metagenomics by increasing sequencing depth without increasing replication. Although we demonstrated a log reduction in the LoQ with a corresponding log increase in sequencing depth (Figure 2B,C), even at ∼100 Gb, the LoQ of quantitative metagenomics remained 2−3 logs above that of ddPCR (∼10 gc/uL). At a more typical sequencing depth of 10 Gb, we estimate an LoQ of 2.5×10^4 gc/μL, or 2.7×10^3 gc/mL in influent wastewater. Practically speaking, these results beget implications for many conventional shotgun metagenomics studies tracking gene abundances in dynamic systems where the absence of a gene may be interpreted as being "removed" by the underlying biophysical processes but may still be present and biologically relevant. This issue is likely exacerbated in environmental AMR studies where the majority of ARGs are present at low abundance, confounding ecological observations and assessment of engineering controls, among other issues.

While this study was designed using relatively exorbitant experimental bounds, it was intended to shed light on possible limitations of current environmental monitoring paradigms using shotgun metagenomics. Utilizing a commercial lab, we

https://doi.org/10.1021/acs.est.4c08284
Environ. Sci. Technol. 2025, 59, 6192−6202

estimate the total cost per sample for this experiment to be ∼ $1000. This is likely impractical both monetarily and computationally for both routine WBS and environmental monitoring efforts in general, where ddPCR and qPCR retain their value when there is a well-characterized indicator gene present at low concentrations.[11] For routine environmental monitoring, shotgun metagenomics may be especially useful as a prescreening tool for empirically deriving the most appropriate ARG targets globally[48] or for specific environments,[49] then designing primers for more sensitive, rapid, and cost-effective quantification.[50] A key objective of environmental AMR monitoring is the curation of high-resolution spatiotemporal data sets that establish baselines and allow the assessment of interventions or ecological perturbations,[18,51] which is likely more practically suited for PCR-based methods at scale. It is noteworthy that the detection thresholds of quantitative metagenomics illuminated here were within range of those reported for HT-qPCR, where the LoQ has been shown to range from 10^1-10^4 gc/μL depending on the primer set and optimization.[52] The limitations of HT-qPCR have been discussed previously, but the approach is a viable option for environmental monitoring of ARGs and other biomarkers given proper advancements in method QA/QC,[53] and has already proven useful in WBS efforts of AMR.[54]

In the context of WBS and broader utility from an epidemiological standpoint, the higher detection thresholds of quantitative metagenomics would likely relegate the approach to high-abundance biomarkers, which may not be suitable for transient diseases that only ever exist at low concentrations in wastewater. The high detection limits and sample processing times also undercut the technique as an early warning system, where low-abundance biomarkers go undetected, which may lead to underreporting or delayed detection of emerging outbreaks. Considering the stochasticity of low-abundance oligonucleotide detection observed in this study (Figures 1C and 2A), the risk of false negatives may outweigh the benefits of nontargeted sequencing for certain etiological agents. For AMR, however, it is important to recognize the utility of shotgun sequencing data for the detection and prediction of novel or latent environmental ARG reservoirs that may be recruited into pathogens.[55−57] Additionally, it is important to also consider broader value to environmental monitoring and assessment where quantitative metagenomic data in time series could provide new insights into ecological processes in natural and engineered systems and the ability to model them. For example, the characterization of canonical metabolic pathways involved in nutrient cycling where subject organisms are enriched and their genes are in high abundance. Recently, culture-enriched phenotypic metagenomic strategies have proven to be a feasible alternative for in-depth and sensitive characterization of beta-lactam resistomes in wastewaters and receiving waters.[58] Similarly, shotgun-sequenced culture enrichments have enhanced the detection of critically important AMR determinants in surface waters as a part of the U.S.'s National Antimicrobial Resistance Monitoring System.[59,60] However, a key limitation of any enrichment technique is the foregoing of the quantitative capacity of the sample, marking a distinct trade-off.

We observed high levels of variance in estimated gene concentrations between replicates using quantitative metagenomics, particularly for beta-lactamase, fluoroquinolone, and tetracycline ARG classes containing several subtypes with arrays of closely related variants. Notwithstanding the biological variability of the replicates, these results highlighted the inherent limitations of short-read Illumina data for "gene counting" where a singular read/read pair can align to multiple variants within set alignment thresholds, but only a single variant can be chosen and counted. For example, the TEM beta-lactamases have 202 unique variants in CARD (v3.0.3) with an average length of 286 amino acids. At 99% sequence identity and 100% coverage, the set reduces to 83 gene clusters, and then to only 27 at 80% (data not shown). Given the limited coverage of a 150 bp (50 amino acids) read, it is an intrinsic limitation of short-read metagenomics to reliably quantify variants with read mapping where a 1:1 read alignment to gene copy ratio is assumed. In general, we found that increasing the specificity of the alignment either converged or diverged on the "true value" that is the ddPCR measurement and thus refer to quantitative metagenomic gene concentrations as estimates. Still, the use of RPK normalized read counts of reference standards to calibrate the "rate" at which reads are mapped to genes was found directly proportional to their absolute abundances, with high fidelity for those at or above the LoQ.

For quantitative metagenomics to be widely and successfully adopted, it will be necessary to establish common protocols for sample processing.[22,61] This includes the spiking of internal reference standards, defining statistical and bioinformatic thresholds for target detection, and benchmarking with PCR-based methods. Here we demonstrated a portion of an ideal quantitative metagenomics workflow, spiking in reference standards to samples after DNA extraction to isolate bias to library preparation and sequencing.[22] We acknowledge that DNA extraction represents one the greatest sources of bias in any metagenomics study, but accounting for extraction efficiency would not have contributed substantially to the present experiment, as it would have only fractionally accounted for lost gene quantities, a bias that was uniformly applied across the data set. Because each extract represents a discrete concentration of genes to be independently challenged, the approach here allowed the direct comparison of quantitative metagenomics to ddPCR measurements and the benchmarking of detection limits. Future quantitative metagenomic studies would benefit from exogenous whole-cell spike-ins to characterize and calibrate individual samples for extraction bias to increase representativeness and accuracy.

The quantitative metagenomics experiment demonstrated here has the potential to broadly expand the capacity of environmental monitoring, not only for ARGs, but also other gene targets and pathogens. For example, marker-based taxonomic annotators such as MetaPhlAn[62] could be used in conjunction with exogenous cell spike-ins (e.g., the Zymo Spike-in controls) for absolute quantitation of individual colonies (i.e., derivation of cfu/mL data). Great strides in quantitative long-read sequencing using exogenous cell spike-ins and cell viability assays have already been made for near-realtime volumetric quantification of both ARGs and pathogens in environmental systems.[63−65] These methods, in conjunction with meta sequin standards should provide the full complement of reference material necessary for comprehensive QA/QC and volumetric quantification of microbial communities for high-throughput sequencing applications, including environmental surveillance, quantitative microbial risk assessment, and microbial source tracking in the future.

https://doi.org/10.1021/acs.est.4c08284
Environ. Sci. Technol. 2025, 59, 6192−6202

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.4c08284.

Ladder fidelity by sample matrix and m/m% (Figure S1), sequin detection by input proportion (Figure S2), ladder viability by m/m% and sequencing depth (Figure S3), results of Kruskal−Wallis tests comparing ladder fidelity across matrix and sequencing depth (Figure S4), summary of reference length and GC bias on sequin detection (Figure S5), sequencing recovery of meta sequin ladders (Figure S6), erroneous sequin alignments (Figure S7), absolute abundance of ARGs by class (Figure S8), LOQ of ARG detection by matrix (Figure S9), log reduction value comparison (Figure S10) (PDF)

ddPCR assay information (Table S1), meta sequin "Mixture A" specifications (Table S2), sample summary statistics (Table S3), summary of log−log models across spike-in m/m% (Table S4), ladder linearity with sequencing depth (Table S5), results of method equivalence testing (Table S6), quantitative metagenomic gene concentrations across all samples (Table S7) (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Benjamin C. Davis − *Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, Ohio 45268, United States; Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0002-8752-6608; Email: davis.benjamin@epa.gov

Authors

Peter J. Vikesland − *Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0003-2654-5132

Amy Pruden − *Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0002-3191-6244

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.4c08284

Notes

The research presented was not performed or funded by EPA and was not subject to EPA's quality system requirements. The views expressed in this article are those of the author(s) and do not necessarily represent the views or the policies of the U.S. Environmental Protection Agency.

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank Gabriel M. Rivera for help with sample processing and ddPCR, Katie Scott for assistance with data analysis, and Scott Keely and Nichole Brinkman for thorough and insightful internal reviews. The authors acknowledge Advanced Research Computing at Virginia Tech for providing computational resources and technical support that have contributed to the results reported within this paper (https://arc.vt.edu/). We acknowledge Water Research Foundation Project 4961 (A.P.), the Centers for Disease Control contract 75D30122C14707 (A.P., P.J.V.), and National Science Foundation NRT Award 2125798 (A.P., P.J.V.) for funding this research.

■ REFERENCES

(1) Mao, K.; Zhang, K.; Du, W.; Ali, W.; Feng, X.; Zhang, H. The potential of wastewater-based epidemiology as surveillance and early warning of infectious disease outbreaks. *Curr. Opin Environ. Sci. Health* **2020**, *17*, 1−7.

(2) Sims, N.; Kasprzyk-Hordern, B. Future perspectives of wastewater-based epidemiology: Monitoring infectious disease spread and resistance to the community level. *Environ. Int.* **2020**, *139*, No. 105689.

(3) Keshaviah, A.; Diamond, M. B.; Wade, M. J.; Scarpino, S. V.; Global Wastewater Action Group. Wastewater monitoring can anchor global disease surveillance systems. *Lancet Glob Health* **2023**, *11* (6), e976−e981.

(4) Alcock, B. P.; Raphenya, A. R.; Lau, T. T. Y.; Tsang, K. K.; Bouchard, M.; Edalatmand, A.; Huynh, W.; Nguyen, A. V.; Cheng, A. A.; Liu, S.; Min, S. Y.; Miroshnichenko, A.; Tran, H. K.; Werfalli, R. E.; Nasir, J. A.; Oloni, M.; Speicher, D. J.; Florescu, A.; Singh, B.; Faltyn, M.; Hernandez-Koutoucheva, A.; Sharma, A. N.; Bordeleau, E.; Pawlowski, A. C.; Zubyk, H. L.; Dooley, D.; Griffiths, E.; Maguire, F.; Winsor, G. L.; Beiko, R. G.; Brinkman, F. S. L.; Hsiao, W. W. L.; Domselaar, G. V.; McArthur, A. G. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **2019**, *48* (D1), D517−D525.

(5) Medema, G.; Heijnen, L.; Elsinga, G.; Italiaander, R.; Brouwer, A. Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands. *Environmental Science & Technology Letters* **2020**, *7* (7), 511−516.

(6) Lodder, W. J.; Buisman, A. M.; Rutjes, S. A.; Heijne, J. C.; Teunis, P. F.; de Roda Husman, A. M. Feasibility of quantitative environmental surveillance in poliovirus eradication strategies. *Appl. Environ. Microbiol.* **2012**, *78* (11), 3800−3805.

(7) Davis, B. C.; Riquelme, M. V.; Ramirez-Toro, G.; Bandaragoda, C.; Garner, E.; Rhoads, W. J.; Vikesland, P.; Pruden, A. Demonstrating an Integrated Antibiotic Resistance Gene Surveillance Approach in Puerto Rican Watersheds Post-Hurricane Maria. *Environ. Sci. Technol.* **2020**, *54* (23), 15108−15119.

(8) Cho, S.; Hiott, L. M.; Read, Q. D.; Damashek, J.; Westrich, J.; Edwards, M.; Seim, R. F.; Glinski, D. A.; Bateman McDonald, J. M.; Ottesen, E. A.; Lipp, E. K.; Henderson, W. M.; Jackson, C. R.; Frye, J. G. Distribution of Antibiotic Resistance in a Mixed-Use Watershed and the Impact of Wastewater Treatment Plants on Antibiotic Resistance in Surface Water. *Antibiotics (Basel)* **2023**, *12* (11), 1586.

(9) Garner, E.; Inyang, M.; Garvey, E.; Parks, J.; Glover, C.; Grimaldi, A.; Dickenson, E.; Sutherland, J.; Salveson, A.; Edwards, M. A.; Pruden, A. Impact of blending for direct potable reuse on premise plumbing microbial ecology and regrowth of opportunistic pathogens and antibiotic resistant bacteria. *Water Res.* **2019**, *151*, 75−86.

(10) Mao, D.; Yu, S.; Rysz, M.; Luo, Y.; Yang, F.; Li, F.; Hou, J.; Mu, Q.; Alvarez, P. J. Prevalence and proliferation of antibiotic resistance genes in two municipal wastewater treatment plants. *Water Res.* **2015**, *85*, 458−466.

(11) Keenum, I.; Liguori, K.; Calarco, J.; Davis, B. C.; Milligan, E.; Harwood, V. J.; Pruden, A. A framework for standardized qPCR-targets and protocols for quantifying antibiotic resistance in surface water, recycled water and wastewater. *Critical Reviews in Environmental Science and Technology* **2022**, *52* (24), 4395−4419.

(12) Waseem, H.; Jameel, S.; Ali, J.; Saleem Ur Rehman, H.; Tauseef, I.; Farooq, U.; Jamal, A.; Ali, M. I. Contributions and Challenges of High Throughput qPCR for Determining Antimicrobial Resistance in the Environment: A Critical Review. *Molecules* **2019**, *24* (1), 163.

(13) Garner, E.; Davis, B. C.; Milligan, E.; Blair, M. F.; Keenum, I.; Maile-Moskowitz, A.; Pan, J.; Gnegy, M.; Liguori, K.; Gupta, S.;

Prussin, A. J., 2nd; Marr, L. C.; Heath, L. S.; Vikesland, P. J.; Zhang, L.; Pruden, A. Next generation sequencing approaches to evaluate water and wastewater quality. *Water Res.* **2021**, *194*, No. 116907.
(14) Hendriksen, R. S.; Munk, P.; Njage, P.; van Bunnik, B.; McNally, L.; Lukjancenko, O.; Roder, T.; Nieuwenhuijse, D.; Pedersen, S. K.; Kjeldgaard, J.; Kaas, R. S.; Clausen, P.; Vogt, J. K.; Leekitcharoenphon, P.; van de Schans, M. G. M.; Zuidema, T.; de Roda Husman, A. M.; Rasmussen, S.; Petersen, B.; Global Sewage Surveillance project consortium; Amid, C.; Cochrane, G.; Sicheritz-Ponten, T.; Schmitt, H.; Alvarez, J. R. M.; Aidara-Kane, A.; Pamp, S. J.; Lund, O.; Hald, T.; Woolhouse, M.; Koopmans, M. P.; Vigre, H.; Petersen, T. N.; Aarestrup, F. M. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat. Commun.* **2019**, *10* (1), 1124.
(15) Majeed, H. J.; Riquelme, M. V.; Davis, B. C.; Gupta, S.; Angeles, L.; Aga, D. S.; Garner, E.; Pruden, A.; Vikesland, P. J. Evaluation of Metagenomic-Enabled Antibiotic Resistance Surveillance at a Conventional Wastewater Treatment Plant. *Front Microbiol* **2021**, *12*, No. 657954.
(16) Riquelme, M. V. P.; Garner, E.; Gupta, S.; Metch, J.; Zhu, N.; Blair, M. F.; Arango-Argoty, G.; Maile-Moskowitz, A.; Li, A. D.; Flach, C. F.; Aga, D. S.; Nambi, I. M.; Larsson, D. G. J.; Burgmann, H.; Zhang, T.; Pruden, A.; Vikesland, P. J. Demonstrating a Comprehensive Wastewater-Based Surveillance Approach That Differentiates Globally Sourced Resistomes. *Environ. Sci. Technol.* **2022**, *56* (21), 14982−14993.
(17) Aarestrup, F. M.; Woolhouse, M. E. J. Using sewage for surveillance of antimicrobial resistance. *Science* **2020**, *367* (6478), 630−632.
(18) Pruden, A.; Vikesland, P. J.; Davis, B. C.; de Roda Husman, A. M. Seizing the moment: now is the time for integrated global surveillance of antimicrobial resistance in wastewater environments. *Curr. Opin Microbiol* **2021**, *64*, 91−99.
(19) Gloor, G. B.; Macklaim, J. M.; Pawlowsky-Glahn, V.; Egozcue, J. J. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front Microbiol* **2017**, *8*, 2224.
(20) Garner, E.; Organiscak, M.; Dieter, L.; Shingleton, C.; Haddix, M.; Joshi, S.; Pruden, A.; Ashbolt, N. J.; Medema, G.; Hamilton, K. A. Towards risk assessment for antibiotic resistant pathogens in recycled water: a systematic review and summary of research needs. *Environ. Microbiol* **2021**, *23* (12), 7355−7372.
(21) Haas, C. N. Quantitative Microbial Risk Assessment and Molecular Biology: Paths to Integration. *Environ. Sci. Technol.* **2020**, *54* (14), 8539−8546.
(22) Davis, B. C.; Brown, C.; Gupta, S.; Calarco, J.; Liguori, K.; Milligan, E.; Harwood, V. J.; Pruden, A.; Keenum, I. Recommendations for the use of metagenomics for routine monitoring of antibiotic resistance in wastewater and impacted aquatic environments. *Critical Reviews in Environmental Science and Technology* **2023**, *53* (19), 1731−1756.
(23) Yin, X.; Chen, X.; Jiang, X. T.; Yang, Y.; Li, B.; Shum, M. H.; Lam, T. T. Y.; Leung, G. M.; Rose, J.; Sanchez-Cid, C.; Vogel, T. M.; Walsh, F.; Berendonk, T. U.; Midega, J.; Uchea, C.; Frigon, D.; Wright, G. D.; Bezuidenhout, C.; Picao, R. C.; Ahammad, S. Z.; Nielsen, P. H.; Hugenholtz, P.; Ashbolt, N. J.; Corno, G.; Fatta-Kassinos, D.; Burgmann, H.; Schmitt, H.; Cha, C. J.; Pruden, A.; Smalla, K.; Cytryn, E.; Zhang, Y.; Yang, M.; Zhu, Y. G.; Dechesne, A.; Smets, B. F.; Graham, D. W.; Gillings, M. R.; Gaze, W. H.; Manaia, C. M.; van Loosdrecht, M. C. M.; Alvarez, P. J. J.; Blaser, M. J.; Tiedje, J. M.; Topp, E.; Zhang, T. Toward a Universal Unit for Quantification of Antibiotic Resistance Genes in Environmental Samples. *Environ. Sci. Technol.* **2023**, *57* (26), 9713−9721.
(24) Hardwick, S. A.; Deveson, I. W.; Mercer, T. R. Reference standards for next-generation sequencing. *Nat. Rev. Genet* **2017**, *18* (8), 473−484.
(25) Hardwick, S. A.; Chen, W. Y.; Wong, T.; Kanakamedala, B. S.; Deveson, I. W.; Ongley, S. E.; Santini, N. S.; Marcellin, E.; Smith, M. A.; Nielsen, L. K.; Lovelock, C. E.; Neilan, B. A.; Mercer, T. R. Synthetic microbe communities provide internal reference standards for metagenome sequencing and analysis. *Nat. Commun.* **2018**, *9* (1), 3096.
(26) Crossette, E.; Gumm, J.; Langenfeld, K.; Raskin, L.; Duhaime, M.; Wigginton, K. Metagenomic Quantification of Genes with Internal Standards. *mBio* **2021**, *12* (1), 20.
(27) Li, B.; Li, X.; Yan, T. A Quantitative Metagenomic Sequencing Approach for High-Throughput Gene Quantification and Demonstration with Antibiotic Resistance Genes. *Appl. Environ. Microbiol.* **2021**, *87* (16), No. e0087121.
(28) Suzuki, M. T.; Taylor, L. T.; DeLong, E. F. Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5′-nuclease assays. *Appl. Environ. Microbiol.* **2000**, *66* (11), 4605−4614.
(29) Pei, R.; Kim, S. C.; Carlson, K. H.; Pruden, A. Effect of river landscape on the sediment concentrations of antibiotics and corresponding antibiotic resistance genes (ARG). *Water Res.* **2006**, *40* (12), 2427−2435.
(30) Barraud, O.; Baclet, M. C.; Denis, F.; Ploy, M. C. Quantitative multiplex real-time PCR for detecting class 1, 2 and 3 integrons. *J. Antimicrob. Chemother.* **2010**, *65* (8), 1642−1645.
(31) Colomer-Lluch, M.; Jofre, J.; Muniesa, M. Antibiotic resistance genes in the bacteriophage DNA fraction of environmental samples. *PLoS One* **2011**, *6* (3), No. e17549.
(32) Dutka-Malen, S.; Evers, S.; Courvalin, P. Detection of Glycopeptide Resistance Genotypes and Identification to the Species Level of Clinically Relevant Enterococci by PCR. *Journal of Clinical Microbiology* **1995**, *33* (1), 24−27.
(33) Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34* (17), i884−i890.
(34) Rodriguez, R. L. M.; Gunturu, S.; Tiedje, J. M.; Cole, J. R.; Konstantinidis, K. T. Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity. *mSystems* **2018**, *3* (3), No. e00039-18.
(35) Wong, T.; Deveson, I. W.; Hardwick, S. A.; Mercer, T. R. ANAQUIN: a software toolkit for the analysis of spike-in controls for next generation sequencing. *Bioinformatics* **2017**, *33* (11), 1723−1724.
(36) DeSantis, T. Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E. L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G. L. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **2006**, *72* (7), 5069−5072.
(37) Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34* (18), 3094−3100.
(38) Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25* (16), 2078−2079.
(39) Shen, W.; Le, S.; Li, Y.; Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One* **2016**, *11* (10), No. e0163962.
(40) Lakens, D. Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses. *Soc. Psychol Personal Sci.* **2017**, *8* (4), 355−362.
(41) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2024. https://www.R-project.org/ (accessed May 24, 2023).
(42) Reis, A. L. M.; Deveson, I. W.; Wong, T.; Madala, B. S.; Barker, C.; Blackburn, J.; Marcellin, E.; Mercer, T. R. A universal and independent synthetic DNA ladder for the quantitative measurement of genomic features. *Nat. Commun.* **2020**, *11* (1), 3609.
(43) Langenfeld, K.; Hegarty, B.; Vidaurri, S.; Crossette, E.; Duhaime, M.; Wigginton, K. A quantitative metagenomic approach to determine population concentrations with examination of quantitative limitations. *bioRxiv* **2022**, No. 499345.
(44) Forootan, A.; Sjoback, R.; Bjorkman, J.; Sjogreen, B.; Linz, L.; Kubista, M. Methods to determine limit of detection and limit of quantification in quantitative real-time PCR (qPCR). *Biomol Detect Quantif* **2017**, *12*, 1−6.

https://doi.org/10.1021/acs.est.4c08284
Environ. Sci. Technol. 2025, 59, 6192−6202

(45) Wang, J.; Chu, L.; Wojnarovits, L.; Takacs, E. Occurrence and fate of antibiotics, antibiotic resistant genes (ARGs) and antibiotic resistant bacteria (ARB) in municipal wastewater treatment plant: An overview. *Sci. Total Environ.* **2020**, *744*, No. 140997.
(46) Wang, D.; Wang, S.; Du, X.; He, Q.; Liu, Y.; Wang, Z.; Feng, K.; Li, Y.; Deng, Y. ddPCR surpasses classical qPCR technology in quantitating bacteria and fungi in the environment. *Molecular Ecology Resources* **2022**, *22*, 2587−2598.
(47) Taylor, S. C.; Laperriere, G.; Germain, H. Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. *Sci. Rep* **2017**, *7* (1), 2409.
(48) Zhang, Z.; Zhang, Q.; Wang, T.; Xu, N.; Lu, T.; Hong, W.; Penuelas, J.; Gillings, M.; Wang, M.; Gao, W.; Qian, H. Assessment of global health risk of antibiotic resistance genes. *Nat. Commun.* **2022**, *13* (1), 1553.
(49) Tarek, M. H.; Garner, E. A proposed framework for the identification of indicator genes for monitoring antibiotic resistance in wastewater: Insights from metagenomic sequencing. *Sci. Total Environ.* **2023**, *854*, No. 158698.
(50) Tarek, M. H.; Hubbart, J. A.; Garner, E. Tracking Sources and Dissemination of Indicator Antibiotic Resistance Genes at a Watershed Scale. *ES&T Water* **2024**, *4* (2), 399−412.
(51) Bengtsson-Palme, J.; Abramova, A.; Berendonk, T. U.; Coelho, L. P.; Forslund, S. K.; Gschwind, R.; Heikinheimo, A.; Jarquin-Diaz, V. H.; Khan, A. A.; Klumper, U.; Lober, U.; Nekoro, M.; Osinska, A. D.; Ugarcina Perovic, S.; Pitkanen, T.; Rodland, E. K.; Ruppe, E.; Wasteson, Y.; Wester, A. L.; Zahra, R. Towards monitoring of antimicrobial resistance in the environment: For what reasons, how to implement it, and what are the data needs? *Environ. Int.* **2023**, *178*, No. 108089.
(52) Crane, S. L.; van Dorst, J.; Hose, G. C.; King, C. K.; Ferrari, B. C. Microfluidic qPCR Enables High Throughput Quantification of Microbial Functional Genes but Requires Strict Curation of Primers. *Front. Environ. Sci.* **2018**, *6*, 145.
(53) Hill, E. R.; Chun, C. L.; Hamilton, K.; Ishii, S. High-Throughput Microfluidic Quantitative PCR Platform for the Simultaneous Quantification of Pathogens, Fecal Indicator Bacteria, and Microbial Source Tracking Markers. *ACS ES T Water* **2023**, *3* (8), 2647−2658.
(54) Pärnänen, K.; Narciso-da-Rocha, C.; Kneis, D.; Berendonk, T.; Cacace, D.; Do, T. T.; Elpers, C.; Fatta-Kassinos, D.; Henriques, I.; Jaeger, T.; Karkman, A.; Martinez, J. L.; Michael, S. G.; Michael-Kordatou, I.; O'Sullivan, K.; Rodriguez-Mozaz, S.; Schwartz, T.; Sheng, H.; Sørum, H.; Stedtfeld, R. D.; Tiedje, J. M.; Della Giustina, S. V.; Walsh, F.; Vaz-Moreira, I.; Virta, M.; Manaia, C. M. Antibiotic resistance in European wastewater treatment plants mirrors the pattern of clinical antibiotic resistance prevalence. *Sci. Adv.* **2019**, *5* (3), No. eaau9124.
(55) Inda-Diaz, J. S.; Lund, D.; Parras-Molto, M.; Johnning, A.; Bengtsson-Palme, J.; Kristiansson, E. Latent antibiotic resistance genes are abundant, diverse, and mobile in human, animal, and environmental microbiomes. *Microbiome* **2023**, *11* (1), 44.
(56) Arango-Argoty, G.; Garner, E.; Pruden, A.; Heath, L. S.; Vikesland, P.; Zhang, L. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* **2018**, *6* (1), 23.
(57) Berglund, F.; Osterlund, T.; Boulund, F.; Marathe, N. P.; Larsson, D. G. J.; Kristiansson, E. Identification and reconstruction of novel antibiotic resistance genes from metagenomes. *Microbiome* **2019**, *7* (1), 52.
(58) Zhang, Z.; Zhang, G.; Ju, F. Using Culture-Enriched Phenotypic Metagenomics for Targeted High-Throughput Monitoring of the Clinically Important Fraction of the beta-Lactam Resistome. *Environ. Sci. Technol.* **2022**, *56* (16), 11429−11439.
(59) Ottesen, A.; Kocurek, B.; Ramachandran, P.; Reed, E.; Commichaux, S.; Engelbach, G.; Mammel, M.; Saint Fleurant, S.; Zhao, S.; Kabera, C.; Merrill, A.; Bonin, N.; Worley, H.; Noyes, N.; Boucher, C.; McDermott, P.; Strain, E. Advancing antimicrobial resistance monitoring in surface waters with metagenomic and quasimetagenomic methods. *PLOS Water* **2022**, *1* (12), No. e0000067.
(60) Franklin, A. M.; Weller, D. E.; Durso, L. M.; Bagley, M. J.; Davis, B. C.; Frye, J. G.; Grim, C. J.; Ibekwe, A. M.; Jahne, M.; Keely, S.; Kraft, A. L.; McConn, B. R.; Mitchell, R. M.; Ottesen, A.; Sharma, M.; Strain, E.; Tadesse, D. A.; Tate, H.; Wells, J. E.; Williams, C. F.; Cook, K. L.; Kabera, C.; McDermott, P. F.; Garland, J. A one health approach for monitoring antimicrobial resistance: developing a national freshwater pilot effort. *Frontiers in Water* **2024**, *6*, 1359109.
(61) Mao, X.; Yin, X.; Yang, Y.; Che, Y.; Xu, X.; Deng, Y.; Li, L.; Zhang, T. Standardization in global environmental antibiotic resistance genes (ARGs) surveillance. *Crit. Rev. Environ. Sci. Technol.* **2024**, *54*, 1633−1650.
(62) Blanco-Miguez, A.; Beghini, F.; Cumbo, F.; McIver, L. J.; Thompson, K. N.; Zolfo, M.; Manghi, P.; Dubois, L.; Huang, K. D.; Thomas, A. M.; Nickols, W. A.; Piccinno, G.; Piperni, E.; Puncochar, M.; Valles-Colomer, M.; Tett, A.; Giordano, F.; Davies, R.; Wolf, J.; Berry, S. E.; Spector, T. D.; Franzosa, E. A.; Pasolli, E.; Asnicar, F.; Huttenhower, C.; Segata, N. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* **2023**, *41* (11), 1633−1644.
(63) Yang, Y.; Deng, Y.; Shi, X.; Liu, L.; Yin, X.; Zhao, W.; Li, S.; Yang, C.; Zhang, T. QMRA of beach water by Nanopore sequencing-based viability-metagenomics absolute quantification. *Water Res.* **2023**, *235*, No. 119858.
(64) Yang, Y.; Che, Y.; Liu, L.; Wang, C.; Yin, X.; Deng, Y.; Yang, C.; Zhang, T. Rapid absolute quantification of pathogens and ARGs by nanopore sequencing. *Sci. Total Environ.* **2022**, *809*, No. 152190.
(65) Wang, C.; Yin, X.; Xu, X.; Wang, D.; Liu, L.; Zhang, X.; Yang, C.; Zhang, X.; Zhang, T. Metagenomic absolute quantification of antibiotic resistance genes and virulence factor genes-carrying bacterial genomes in anaerobic digesters. *Water Res.* **2024**, *253*, No. 121258.

https://doi.org/10.1021/acs.est.4c08284
Environ. Sci. Technol. 2025, 59, 6192−6202