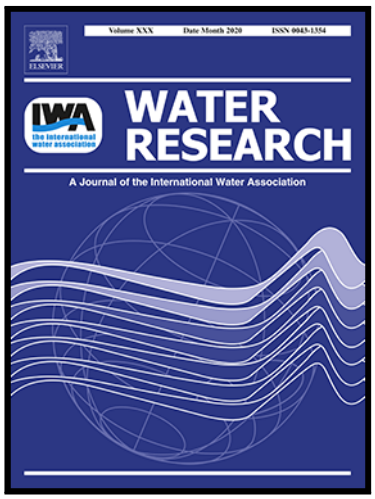



# National-scale antimicrobial resistance surveillance in wastewater: A comparative analysis of HT qPCR and metagenomic approaches

Margaret E. Knight , Gordon Webster , William B. Perry , Amy Baldwin , Laura Rushton , Daniel A. Pass , Gareth Cross , Isabelle Durance , Windi Muziasari , Peter Kille , Kata Farkas , Andrew J. Weightman , Davey L. Jones

PII: S0043-1354(24)00882-0
DOI: https://doi.org/10.1016/j.watres.2024.121989
Reference: WR 121989

To appear in: *Water Research*

Received date: 1 March 2024
Revised date: 18 June 2024
Accepted date: 19 June 2024

Please cite this article as: Margaret E. Knight , Gordon Webster , William B. Perry , Amy Baldwin , Laura Rushton , Daniel A. Pass , Gareth Cross , Isabelle Durance , Windi Muziasari , Peter Kille , Kata Farkas , Andrew J. Weightman , Davey L. Jones , National-scale antimicrobial resistance surveillance in wastewater: A comparative analysis of HT qPCR and metagenomic approaches, *Water Research* (2024), doi: https://doi.org/10.1016/j.watres.2024.121989

**Highlights**

- HT qPCR detected all 73 targeted ARGs, while metagenomics detected 491 ARGs
- HT qPCR was more sensitive to low abundance genes, detecting all target ARGs.
- Both methods enabled the spatiotemporal separation of hospital and WWTP resistomes.
- Metagenomics provided contextual data making it more suitable for risk assessment.
- HT qPCR permitted more sensitive quantification of clinically relevant AMR genes.

# National-scale antimicrobial resistance surveillance in wastewater: A comparative analysis of HT qPCR and metagenomic approaches

Margaret E. Knight[a]*, Gordon Webster[b], William B. Perry[c], Amy Baldwin[b], Laura Rushton[b], Daniel A. Pass[d], Gareth Cross[e], Isabelle Durance[c], Windi Muziasari[f], Peter Kille[b], Kata Farkas[a], Andrew J. Weightman[b], Davey L. Jones[a]

[a] School of Environmental & Natural Sciences, Bangor University, Bangor, LL57 2UW, Wales, UK

[b] Microbiomes, Microbes and Informatics Group, Organisms and Environment Division, School of Biosciences, Museum Avenue, Cardiff University, Cardiff, CF10 3AX, Wales, UK

[c] Water Research Institute, School of Biosciences, Cardiff University, Cardiff, CF10 3AX, Wales, UK

[d] Compass Bioinformatics, 17 Habershon Street, Cardif, CF24 2DU, Wales, UK

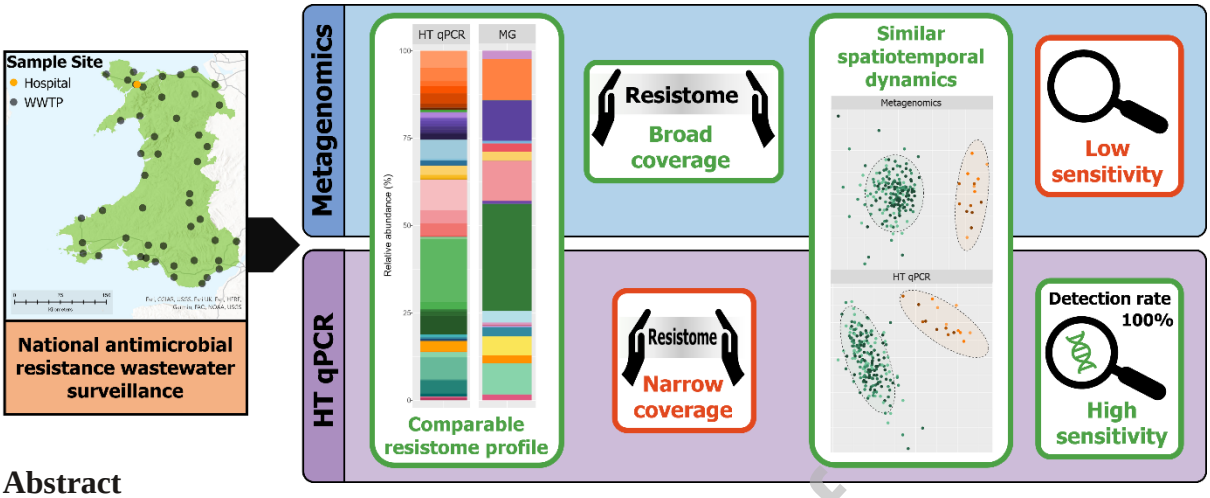
[e] Science Evidence Advice Division, Health and Social Services Group, Welsh Government, Cathays Park, Cardiff, CF10 3NQ, Wales, UK

[f] Resistomap Oy, Cultivator II, Viikinkaari 4, Helsinki, Finland

[*]Corresponding author at: School of Environmental and Natural Sciences, Bangor University, Bangor, LL57 2UW, Wales, UK

*E-mail address*: maggie.knight@bangor.ac.uk (M. Knight).

***Graphical abstract***





## Abstract

Wastewater serves as an important reservoir of antimicrobial resistance (AMR), and its surveillance can provide insights into population-level trends in AMR to inform public health policy. This study compared two common high-throughput screening approaches, namely (i) high-throughput quantitative PCR (HT qPCR), targeting 73 antimicrobial resistance genes, and (ii) metagenomic sequencing. Weekly composite samples of wastewater influent were taken from 47 wastewater treatment plants (WWTPs) across Wales, as part of a national AMR surveillance programme, alongside 4 weeks of daily wastewater effluent samples from a large municipal hospital. Metagenomic analysis provided more comprehensive resistome coverage, detecting 545 genes compared to the targeted 73 genes by HT qPCR. It further provided contextual information critical to risk assessment (i.e. potential bacterial hosts). In contrast, HT qPCR exhibited higher sensitivity, quantifying all targeted genes including those of clinical relevance present at low abundance. When limited to the HT qPCR target genes, both methods were able to reflect the spatiotemporal dynamics of the complete metagenomic resistome, distinguishing that of the hospital and the WWTPs. Both approaches revealed correlations between resistome compositional shifts and environmental variables like ammonium wastewater concentration, though differed in their interpretation of some potential influencing factors. Overall, metagenomics provides more comprehensive resistome

profiling, while qPCR permits sensitive quantification of genes significant to clinical resistance. We highlight the importance of selecting appropriate methodologies aligned to surveillance aims to guide the development of effective wastewater-based AMR monitoring programmes.

*Keywords*: ARG; Environmental reservoir; One Health approach; Wastewater-based epidemiology; WBE strategy.

## 1 Introduction

Antimicrobial resistance (AMR) is a complex problem whose impact traverses human, animal and environmental sectors. Reflecting its multifactorial nature, a greater emphasis has been placed on addressing the problem under the One Health perspective, in recognition that these sectors are fundamentally interconnected (McEwen, Collignon 2018). Urban wastewater represents a significant source of AMR pollution in the environment, but can offer a relatively unbiased estimate of antimicrobial resistance genes (ARGs) circulating in the population (Lai, Muziasari et al. 2021a, Munk, Brinch et al. 2022). As such, it presents a valuable means by which spatiotemporal patterns of AMR can be monitored, avoiding the practical and legal issues associated with public health surveillance when sampling directly from the population (Pruden, Vikesland et al. 2021). Since the predominant fraction of wastewater microbiota is cannot be cultivated by standard methods (Steen, Crits-Christoph et al. 2019), AMR surveillance studies increasingly rely on cultivation-independent methods, primarily quantitative PCR (qPCR) and metagenomic sequencing. Using target-specific primers, qPCR permits the detection and quantification of individual ARGs with prior known sequences. High-throughput qPCR (HT-qPCR) enables the screening of hundreds of such genes in parallel and has been used in numerous studies of AMR prevalence in wastewater and effluent-receiving environments (Karkman, Johnson et al. 2016, Lai, Muziasari et al.

2021a). Metagenomics involves sequencing the genetic material from a sample. Subsequent alignment of DNA reads against a reference database provides a compositional overview of the ARGs within wastewater microbial communities (Brinch, Leekitcharoenphon et al. 2020, Munk, Brinch et al. 2022).

While national wastewater monitoring can further our understanding of AMR, the value of the resultant data depends on aligning a surveillance program's objective with the most suitable approach (Robins, Leonard et al. 2022, Huijbers, Flach et al. 2019). Metagenomics being non-targeted offers the distinct advantage of providing broader coverage of the ARG profile, as previous studies have demonstrated (Ferreira, Otani et al. 2023, Liu, X., Xiao et al. 2019). In contrast, conventional qPCR is typically more sensitive to low abundance genes (Ferreira, Otani et al. 2023); however, the reduced volumes required for high throughput analysis can decrease the method's sensitivity (Stedtfeld, Guo et al. 2018). Overall, studies offering a comprehensive comparison regarding the merits, drawbacks and degree of concordance between these approaches remain limited, especially in the context of wastewater monitoring.

The current study aimed to discern nationwide spatial and temporal variations in the AMR profile of wastewater, and further assess the appropriateness of HT qPCR and NovaSeq metagenomics for different surveillance objectives. To this end, weekly composite influent samples were collected from 47 Welsh wastewater treatment plants (WWTPs) and further effluent samples were taken from a large municipal hospital (Figure 1). The two methods were compared for their gene coverage, their capacity to monitor high risk ARGs —defined as those associated with mobile genetic elements (MGEs), pathogenic hosts, and anthropogenic impacts—and their ability to capture shifts in the resistome composition in relation to environmental variables and the microbial community.

## 2 Material and methods

### *2.1 Wastewater sample collection and processing*

Samples were collected as part of the Welsh Government sponsored wastewater-based public health surveillance programme which has been in operation since September 2020. The programme collects influent wastewater from 47 urban WWTPs within Wales, capturing ca. 66% of population connected to an urban sewer system (Fig. 1). The sites are operated by either Dŵr Cymru/Welsh Water or Hafren Dyfrdwy Cyfyngedig water companies. All sites are equipped with either S320H Aquacell (Aquamatic Ltd., Manchester, UK) or Bühler 4011 (Hach UK Ltd, Manchester, UK) refrigerated autosamplers, positioned after the primary inlet screen which take samples every 15 min to form a daily composite. Samples are taken daily from Monday-Friday and the chilled samples couriered each day to processing laboratories located in Bangor and Cardiff. All WWTP samples were processed within 24 h and combined into weekly composite samples. In addition, untreated effluent samples were collected daily (Tuesday-Friday) from the main drain exiting one large municipal hospital (Ysbyty Gwynedd; Fig. 1) using an ISCO refrigerated autosampler. Both the WWTP and hospital sampling campaigns were conducted over a span of four weeks from March to April 2022. Upon arrival at the processing laboratories, all samples (50 ml aliquots) were centrifuged at 10,000 x g for 30 min at 4°C, and the supernatant discarded. Wastewater biomass pellets were then pooled for each of the individual 47 WWTPs to represent weekly composite samples and suspended in 2 ml sodium phosphate buffer (MP Biomedicals), whereas the daily hospital samples were resuspended directly in 1 ml sodium phosphate buffer and stored at -20°C until required. For details on the measurement of physicochemical and anthropogenic variables, please see the Supplementary Information.

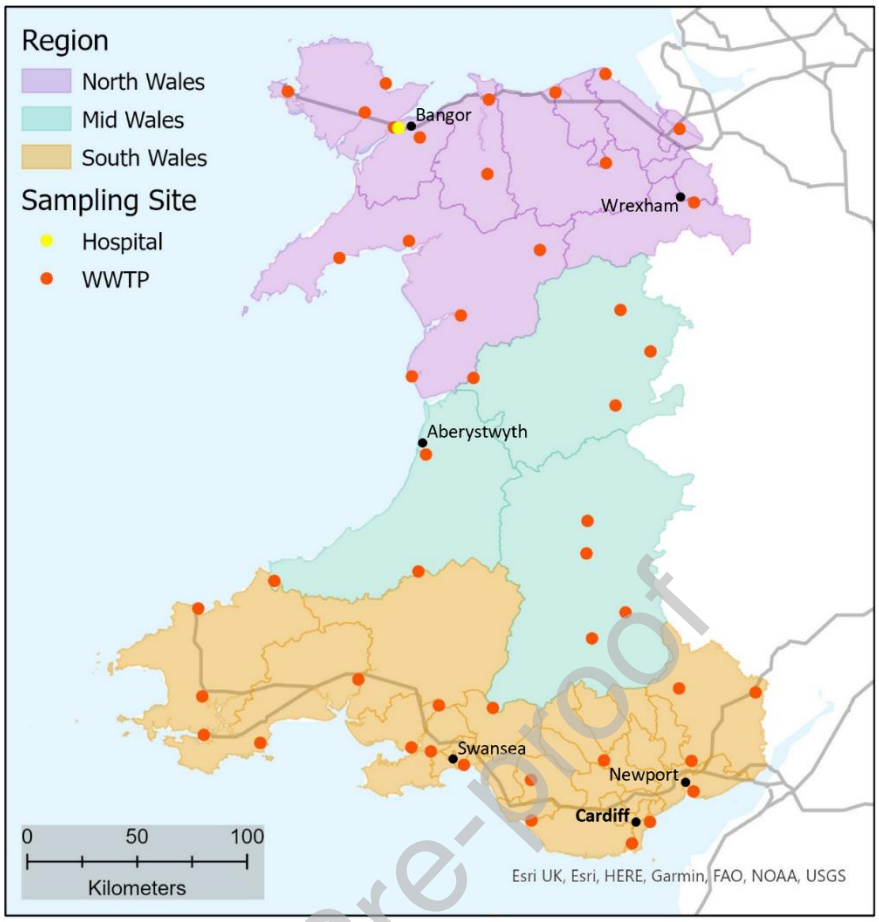



**Figure 1** Map showing the 47 urban wastewater treatment plants and hospital used for sample collection.

*2.2 Wastewater DNA extraction*

Genomic DNA was extracted from wastewater samples (daily hospital or weekly WWTPs) using a modified FastDNA™ SPIN Kit for Soil (MP biomedicals) protocol used routinely for environmental samples (see the Supplementary Information for details) (Webster, Newberry et al. 2003). DNA samples were then stored at -20°C until required.

*2.3 High-throughput qPCR*

A total of 96 primer sets were used in the HT qPCR assay, targeting ARGs (73 genes), MGEs (10 genes), pathogens (5 genes), and 8 other genes including those for metal and biocide resistance. These targets were chosen based on a preliminary screening with 384 genes and advice from the Specialist Antimicrobial Chemotherapy Unit at Public Health Wales to cover

a wide range of clinically relevant ARGs. They were also verified in the MEGAres v3 database for comparison with the metagenomic dataset. The 16S rRNA gene was also quantified to normalise the gene abundances. Aliquots of the DNA extracts were sent to Resistomap Oy (Helsinki, Finland) for gene quantification using the SmartChip™ Real-Time PCR system (TakaraBio, CA, USA). The qPCR cycling conditions and raw data processing have been previously described (Lai, Muziasari et al. 2021b, Muziasari, Pitkanen et al. 2017). Briefly, an initial 10 min denaturation step at 95 °C was followed by 40 cycles of 30 s denaturation at 95 °C and 30 s annealing at 60 °C. All primer sets underwent melting curve analysis; amplicons with non-specific melting curves and multiple peaks were discarded from further analysis. A threshold cycle value ($C_T$) of 27 was chosen as the limit of detection (Lai, Muziasari et al. 2021b, Muziasari, Pitkanen et al. 2017). Each DNA sample was analysed in triplicate; given a gene was detected in at least two technical replicates, the mean $C_T$ was calculated.

### *2.4 Metagenomic library construction and sequencing*

Extracted DNA samples were diluted 1/10 in RSB resuspension buffer (Illumina) and quantified using a Qubit 4 Fluorometer with a Qubit High Sensitivity dsDNA Assay Kit (ThermoFisher Scientific). Library preparation was then carried out using the standard input workflow on 200 ng DNA per sample using the Illumina DNA PCR-Free Prep, Tagmentation Kit and IDT for Illumina DNA/RNA UD Indexes (Sets A-D) as described in the Illumina reference guide. Dual-indexed paired-end single-stranded DNA Libraries (450 bp average library size) were then quantified using the CollibriTM Library Quantification kit (ThermoFisher Scientific) on a LightCycler® 96 Instrument (Roche Diagnostics Ltd.), adjusted to 2.0 nM and pooled. The pooled library was further checked by sequencing on a MiSeq 300 cycle cartridge (Illumina MiSeq System) and the library adjusted to 1.5 nm based

on sequence reads before deep sequencing on an Illumina NovaSeq 6000 Sequencing System (paired end 2 x 150 bp flowcell) with NovaSeq S4 cartridge v1.5 (300 cycles) and 1% PhiX control at Wales Gene Park (www.walesgenepark.cardiff.ac.uk/).

### *2.5 Bioinformatic analysis of sequence data*

Metagenomic sequence reads were converted to FASTQ files from base call files (BCL), demultiplexed and adapter trimmed using Illumina bcl2fastq2 conversion software v2.20. Sequence reads were quality filtered using fastp v 0.20 and checked with fastqc v 0.11.8 and compiled into a single report using MultiQC v1.9 (Ewels, Magnusson et al. 2016). Reads representing host (human) DNA were removed from all sequence files by mapping against the human genome using Samtools v 1.15.1 and Bowtie v2.3.5. High quality non host sequence reads were analysed for ARGs using the AMR++ v3.05 bioinformatic pipeline implemented with NextFlow v21.10. Essentially, reads were aligned to the MEGARes v3 database and ARGs with at least 80% nucleotide coverage were considered present. ARGs requiring SNP confirmation were verified by the SNP confirmation tool. For taxonomic classification reads were also mapped against Kraken2 v2.1.2 (Wood, Lu et al. 2019).

### *2.6 Statistical analysis*

ARG abundances were normalised by the recommended methods for each approach. The metagenomic data was expressed as Fragments Per Kilobase reference per Million bacterial fragments (Munk, Brinch et al. 2022). The HT qPCR data was expressed as ARG copy number /16S rRNA gene copy number (Lai, Muziasari et al. 2021b, Muziasari, Pitkanen et al. 2017). All downstream statistical analyses and figure design was performed on R 4.2.1 (R Core Team 2022) using the vegan, phyloseq, ggplot2 and igraph packages. ARG richness was calculated on count matrices rarefied to the lowest sampling depth (12132 reads). The

number of unique ARGs was calculated using non-rarefied count data. NMDS and dbRDA were calculated based on Bray–Curtis dissimilarity matrices of Hellinger-transformed ARG relative abundances. The statistical difference between two groups was determined by a Welch's two-sample t test, a value of $p < 0.05$ was considered significant. Network analysis was performed on Spearman's rank correlation coefficient values between the relative abundances of ARGs and bacteria genera. Strong, significant correlations ($\rho > 0.7$, $p < 0.01$) were visualised on Gephi 0.10.1 (Bastian, Heymann et al. 2009). Wastewater flow estimates used as an environmental variable in the diversity analyses was calculated using the methodology outlined by (Wilde, Perry et al. 2022).

a unique ARG in HT qPCR refers to each targeted gene individually, while in metagenomics, unique ARGs are defined at the gene group level where sequences are clustered based on similarity.

# 3 Results

## *3.1 Influence of high throughput approaches on quantifying the number and abundance of ARGs in wastewater*

For the NovaSeq metagenomic data analysis, the AMR++ v3.0 pipeline assigned an average 252,599 reads per sample to resistance genes, 79% of which indicated resistance against antimicrobials. Aggregated by group, an average of 181 and 227 individual ARGs were observed in WWTP influent and hospital effluent, respectively (Figure 2.A). In total, 491 unique ARGs were detected across all samples, 61 of which were also detected by HT qPCR (Figure 2.C). A unique ARG in HT qPCR refers to each individual target gene, while in metagenomics, they are defined as uniquely named ARGs at the "group" level of the MEGARes classification scheme (Bonin, Doster et al. 2023). Though HT qPCR was limited to the 73 target ARGs, both methods concurred in detecting significantly higher richness in the hospital effluent samples (Welch's t test: $p < 0.01$ for both datasets). Despite this,

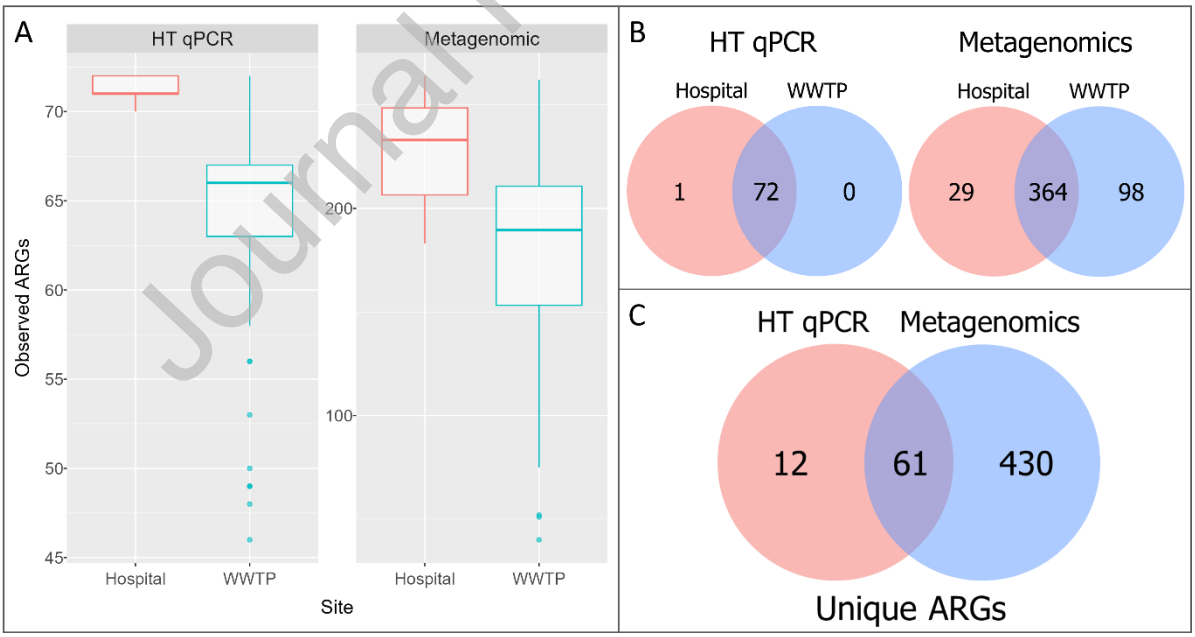



**Figure 2** (A) Box plots displaying the distribution of ARG richness as determined by HT qPCR and metagenomics (note different y-axis scales) across all hospital and WWTP samples. (B) Venn diagrams showing the number of unique ARGs detected in hospital vs WWTP wastewater as determined by HT qPCR and metagenomics. (C) Venn diagram showing the total number of unique ARGs detected in all wastewater samples by HT qPCR vs metagenomics. Note that only 73 ARGs were targeted by HT qPCR.

metagenomic analysis detected a higher number of unique ARGs in the WWTP influent than the hospital effluent (Figure 2.B), likely on account of the greater number of WWTP sampling sites.

The total relative abundance of ARGs was significantly lower in the WWTP influent samples compared to the hospital effluent, whether analysing either dataset (Welch's t-test $p < 0.01$). ARGs for macrolide-lincosamide-streptogramin (MLS), tetracycline, beta lactam and aminoglycoside antimicrobial classes were dominant for both approaches (Figure 3.A and B). They were also concordant in observing a significantly greater relative abundance of glycopeptide resistant ARGs in the hospital effluent (Welch's t-test $p < 0.01$). However, the metagenomic approach further revealed high abundances of ARGs for rifampicin (rifampin) and elfamycin antimicrobial classes, which were not targeted by HT qPCR.

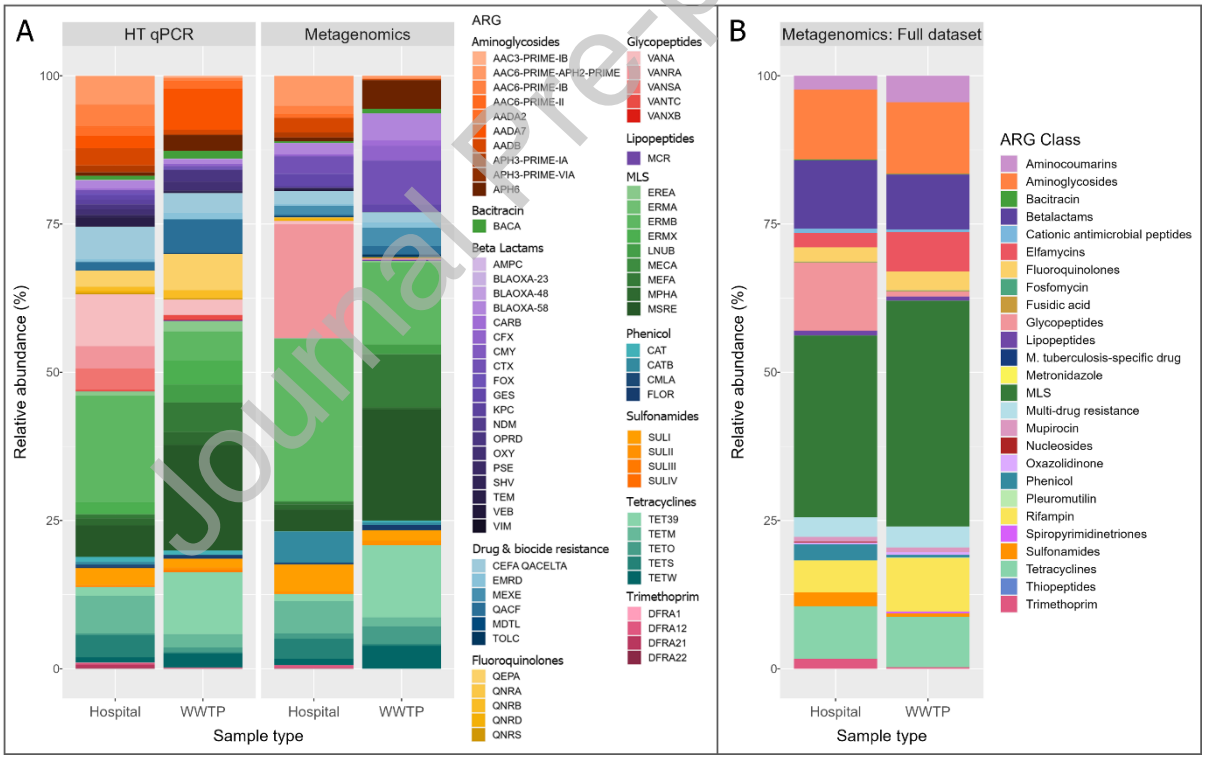



**Figure 3** Normalised relative abundances of (A) the targeted group of ARGs, as determined by HT qPCR and metagenomics and (B) the ARG classes in the full metagenomic dataset, averaged across all WWTPs and all timepoints for both wastewater sources.

Within the scope of the 73 HT qPCR target ARGs, only *mecA* was undetected across the wastewater samples by HT qPCR (Figure 2.B). In comparison, metagenomics detected only

54 and 58 of the HT qPCR target ARGs in WWTP influent and hospital effluent, respectively. Amongst those undetected were high risk ARGs, such as *mdtL*, $bla_{NDM}$ and $bla_{VIM}$ (Zhang, A., Gaston et al. 2021). Detection frequencies remained consistently low across the dataset, such that only 35 of the 73 targeted ARGs were observed in at least 20% of wastewater samples overall. Of these, 22 ARGs exhibited high correlations between the two analysis approaches (Spearman's correlation coefficient $r > 0.7$, $p < 0.01$; Fig. S1). Consequently, when comparing ARG profiles (Figure 3.A), while certain genes (such as *ermB* and *mrsE, tet39* and *tetM*) exhibited similarly high relative abundances and strong positive relationships across both approaches, others differed significantly. ARGs including *aadA7, qepA* and *vanA,* were observed at higher relative abundances by HT qPCR, whereas *tetW, blaCTX* and *vanRA* were more abundant by metagenomic analysis.

Between sample types, the approaches were congruent in finding a significant increase in the relative abundance of *msrE* and *tet39* in the WWTP influent compared to the hospital effluent, and converse was seen for *ermB*. The metagenomic approach further found a significantly greater relative abundance of *ctx* in the WWTP samples, and *vanRA* in the hospital samples. HT qPCR, in contrast, detected a significantly greater relative abundance of *aadA7* in the hospital samples, and *vanA* in the WWTP samples.

### *3.2 Influence of high throughput approaches on evaluating ARG compositional shifts in wastewater*

NDMS was used to compare ARG compositions in the hospital and WWTP wastewater samples over time, revealing resistomes to cluster distinctly by site, whether profiled by either approach (Figure 4) (PERMANOVA $p < 0.01$ for all comparisons). This clustering was influenced by dispersion for the metagenomic dataset, though only when limited to the targeted ARG profile (PERMDISP $p < 0.01$). Assessing the effect of temporal change,

sampling week displayed no discernible impact on the ARG composition in the WWTP influent samples. In contrast, assessment of the NMDS plots indicated an apparent compositional shift across the month in the hospital effluent samples. The temporal variation was most apparent for the metagenomic analysis on the targeted dataset, with genes such as *tetO* and *tetS* increasing in relative abundance in the latter half of the month (Fig. S2).

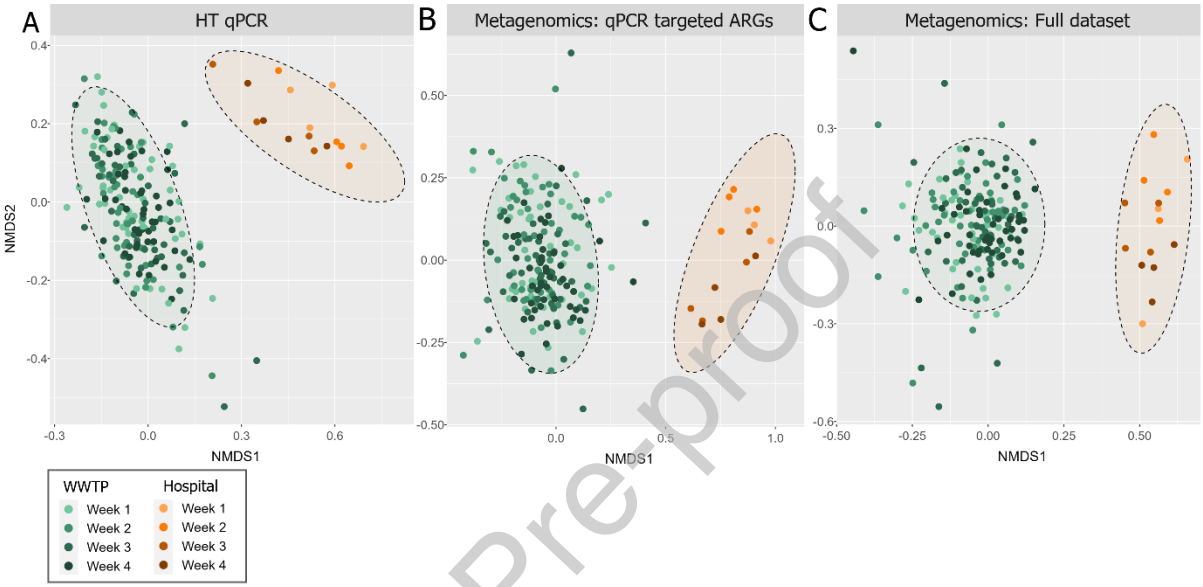



**Figure 4** NMDS ordination plots of Bray–Curtis dissimilarity matrices based on Hellinger-transformed ARG relative abundances, showing ARG composition differences over time and between sampling Hospital and WWTP sampling sites. Determined by (A) HT qPCR and metagenomics based on the (B) qPCR targeted ARGs and (C) full dataset.

### *3.3 Influence of high throughput approaches on determining the impact of environmental and anthropogenic factors on the wastewater resistome*

Assessing the impact of sampling site on the compositional profile of the WWTP influent resistome, NMDS analysis found composition to cluster by geographical region (Fig. S3), as defined in Figure 1 (PERMANOVA $p < 0.01$ for all comparisons). This effect was explained, to some degree, by dispersion (PERMDISP $p < 0.01$). Of the three regions, the resistomes from North and South Wales were the most distinct from one another. The total abundance of ARGs relative to the 16S rRNA gene copy number also differed by sampling site (Fig. S4), such that the average total relative abundance of ARGs, as determined using the HT qPCR

dataset, was significantly lower in South Wales compared to North Wales (Welch's t-test $p < 0.01$).

Distance-based redundancy analysis (dbRDA) revealed that a significant, albeit modest, overall explanatory effect was issued by the anthropogenic and environmental variables on the WWTP influent resistome, with an adjusted $R^2$ of 0.285 ($p < 0.01$) and 0.207 ($p < 0.01$) for HT qPCR and metagenomics, respectively (Figure 5). The two approaches concordantly found the environmental variables to explain a greater percentage of variation in the ARG profile, with ammonium having contributed the largest influence ($R^2 = 0.182$, $p < 0.01$ and $R^2 = 0.102$ $p < 0.01$ for HT qPCR and metagenomics, respectively). There were, however, some minor discrepancies between results concerning the impact of certain variables. For example, the percentage of variation explained by electrical conductivity was found to be significant by metagenomic analysis on the full dataset (Table S1). This significance was, however, not observed in the analysis of the HT qPCR data ($p = 0.359$). Regarding the anthropogenic variables, while their individual influences were significant across the three datasets, the percentage of variation explained by each variable was only low (Table S1).

### *3.4 Influence of high throughput approaches on the relationship of ARGs and bacterial genera*

The composition of the microbial community reflected that of the resistome, being distinct between the WWTP and hospital wastewater (PERMANOVA $p < 0.01$, PERMDISP $p > 0.05$) and shifting over time in the latter of the two sample types (Fig. S5). Pseudomonadota the dominant phylum across all samples, boasting the three most abundant genera, *Acidovorax*, *Acinetobacter* and *Pseudomonas* (Fig. S6).

As an indication for potential host bacteria, co-occurrence patterns between bacterial genera and ARGs were explored by network analysis, based on strong ($\rho > 0.7$) and significant ($p < 0.01$) correlations (Figure 6. A and B). Of the ARGs explored, those indicating resistance to

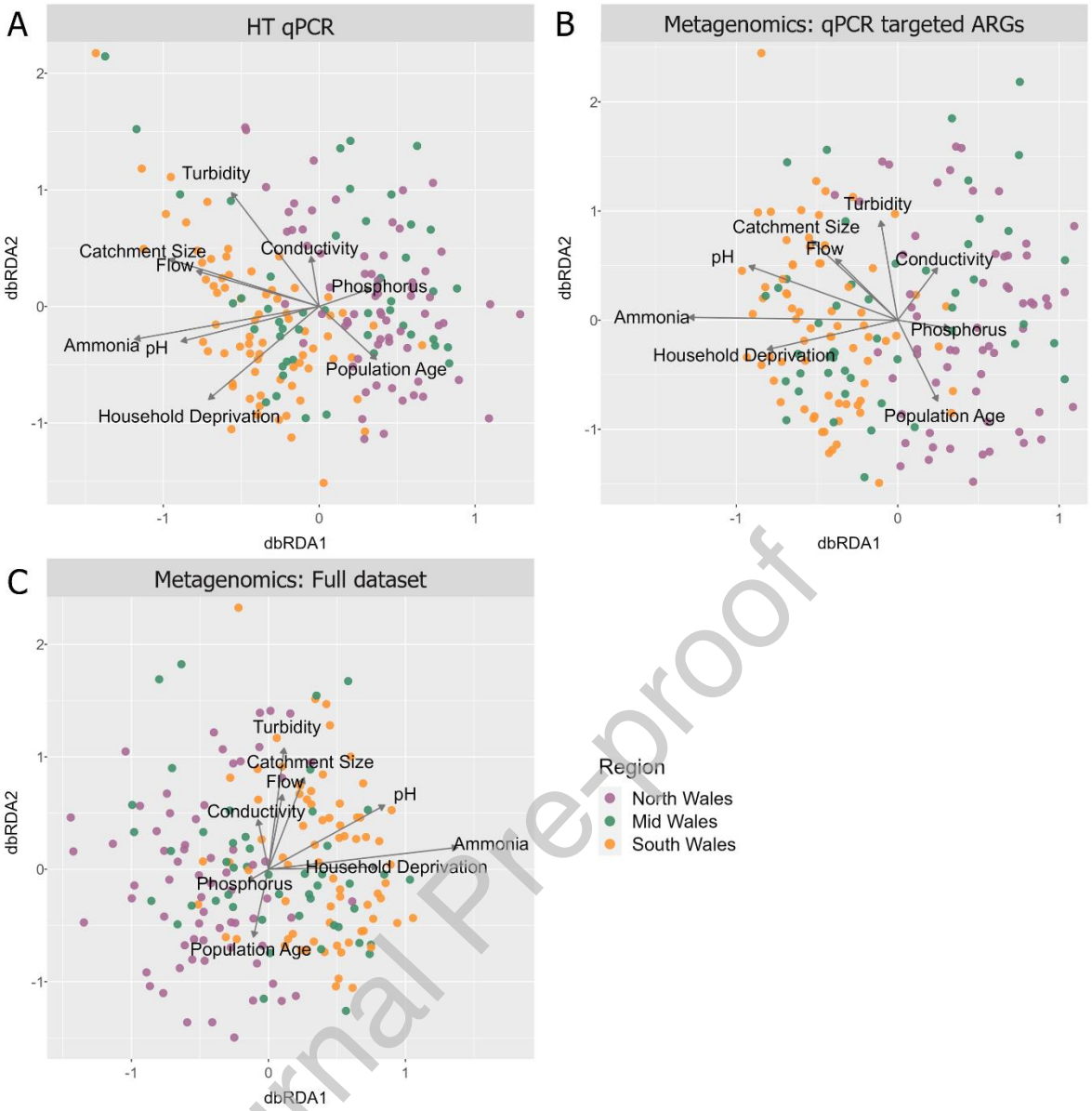



**Figure 5** Distance-based redundancy analysis (dbRDA) showing the influence of environmental and anthropogenic variables on the ARG profile in WWTP influent samples. Determined by (A) HT qPCR and metagenomics based on the (B) qPCR targeted ARGs and (C) full dataset.

tetracyclines displayed the highest number of correlations with different genera. For example, both approaches found *tetO* and *tetW* to exhibit especially strong relationships ($\rho > 0.9$) with two highly abundant Bacillota genera, *Blautia* and *Faecalibacterium*. Metagenomics identified further correlations which HT qPCR did not, including relationships between Pseudomonas and the MLS ARGs *mefA* and *msrD* (Figure 5.B). However, it is also worth noting that the majority of abundant ARGs did not display any significant correlations with any bacterial genera, whether analysing the dataset from either approach.

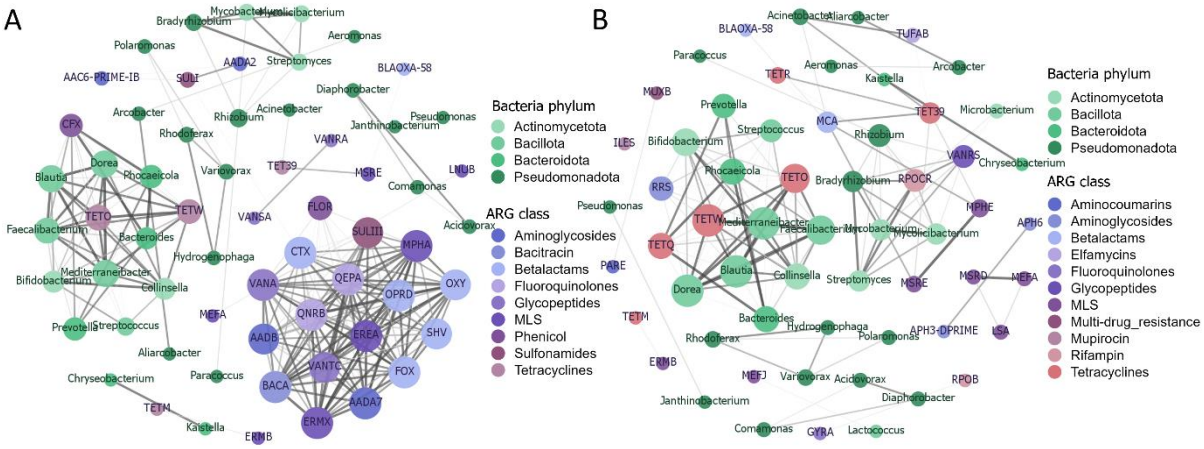



**Figure 6** Network analysis showing the relationships between the top 40 most abundant ARGs and the top 20 most abundant bacterial genera, determined by (A) HT qPCR and (B) Metagenomics. Connections represent strong, significant correlations (Spearman's correlation coefficient $\rho > 0.7$, $p < 0.01$).

## 4 Discussion

AMR wastewater surveillance programs have great potential to monitor emerging health threats, inform public health policy development and support future research. The current study presents a comprehensive comparison of two high throughput approaches which can be applied in such initiatives, as summarised in Table 1.

The results demonstrate that the key factor underpinning the differences between the two methodologies is their distinction in being targeted and untargeted approaches. HT qPCR demonstrated higher sensitivity, detecting all target ARGs, including those present at low abundance. Amongst those were the mobile metallo-β-lactamase, $bla_{\mathrm{NDM}}$ and the multi-drug resistance gene, *mdtL*, both recognised to pose a high risk to human health (Zhang, A., Gaston et al. 2021). In the context of wastewater surveillance, the ability of HT qPCR to monitor genes at low abundances would aid the early detection of outbreaks involving established high-risk ARGs and ensures the continuity of data for routine monitoring. Previous studies on wastewater and manure also found conventional PCR better distinguished smaller deviations in ARG relative abundance patterns (Ferreira, Otani et al. 2023, Crossette, Gumm et al. 2021). However, these studies only assessed a limited number of ARGs and thus could not compare this quality at a compositional level. Here, the results demonstrated that HT qPCR, with enough targets, can effectively reflect the compositional dynamics of the full ARG profile seen by metagenomics. Both approaches distinguished the wastewater influent resistome from that of the hospital effluent. When limited to the same target ARGs, both also captured temporal shifts in the hospital effluent resistome, whereas this effect was slightly obscured when the metagenomic data was broadened to the full dataset. This underlines the importance of determining how best to handle the volume of data generated by metagenomic sequencing. It requires the considered design of appropriate analytical strategies to ensure the

extraction of meaningful results. The data generated by HT qPCR, by comparison, requires a comparatively lower level of expertise to interpret.

On the other hand, given HT qPCR was limited to the predetermined set of 73 genes, metagenomics better captured the diversity within the wastewater resistome. It's worth noting that each qPCR SmartChip has the capacity to measure up to 384 genes; however, a higher number of targets translates to fewer samples which can be analysed on each chip. This greatly increases the cost per sample, making it prohibitive to large scale routine surveillance. Unrestrained to predefined targets, metagenomics is thus better suited to evaluating the overall ARG burden. This could apply to, for example, studies aiming to assess the impact of wastewater release on the diversity and composition of the environmental resistome. It should be acknowledged that gene coverage was also restricted for the metagenomic analysis, though to a far lesser degree, constrained within the scope of the MEGARes database. However, new bioinformatic tools, such as machine learning models (Behling, Wilson et al. 2023) could allow the dataset to be mined for novel ARGs. Furthermore, upon the future discovery of novel ARGs, the data could undergo retrospective analysis to reveal historical trends and evolutionary insights. PCR primer specificity has the additional disadvantage of imposing a bias in ARG quantification (Wei, Feng et al. 2018). This may be further confounded by the running of all PCR reactions in parallel, as some primers might not have performed under optimal conditions. Though to what degree this impacts the accuracy of quantification has not been well explored.

Neither method found the composition of the WWTP influent resistome to vary considerably over the course of the sampling period (ca. one month). In concordance, past longitudinal studies investigating the resistome of influent and activated sludge found the composition to remain relatively stable over time, subject to little seasonal change (Zheng, Huyan et al. 2020, Brinch, Leekitcharoenphon et al. 2020). However, at the level of individual ARGs, temporal

**Table 1** A table comparing the advantages (Green / +), disadvantages (Red / -) and marginal differences (Amber / ▫) of HT qPCR and metagenomics for monitoring AMR in wastewater.

| | HT qPCR | Metagenomics |
|---|---|---|
| **Sensitivity** | + High sensitivity | - Lower sensitivity<br>- Requires spike-ins to establish LOD<br>+ Can be more sensitive to divergent genes |
| **Bioinformatic requirements** | + Data storage and analysis require minimal computational resources<br>+ Data analysis requires a lower bioinformatics skill level | - Datasets and analysis require high (TB) storage and computational power<br>- Data analysis requires high bioinformatic expertise |
| **Quantification** | ▫ Mostly used to report relative abundance data<br>▫ Absolute abundance can be derived from normalisation to the 16S rRNA gene, but introduces bias<br>+ The standard curve method permits absolute quantification | ▫ Produces relative abundance data<br>- Auxiliary quantification of the 16S rRNA gene allows derivation of absolute abundance, but introduces bias<br>- Internal standard spike-ins enables derivation of absolute abundance |
| **Cost** | + Lower cost per sample<br>- The cost per sample increase with the number of target ARGs | - Higher cost per sample, but is falling with technological advancements<br>- Higher read coverage increases cost<br>+ Lower cost per ARG<br>+ Higher cost effectiveness when monitoring large sample numbers |
| **Turnaround time** | ▫ Sample preparation: DNA extraction and PCR setup<br>+ Can process many samples in parallel<br>+ Data analysis is relatively simple and fast | ▫ Sample preparation: DNA extraction, library construction and sequencing preparation<br>+ Can sequence many samples in parallel<br>- Data analysis requires extensive bioinformatic processing |
| **Quantification bias** | - Primer specificity imposes biases<br>- Simultaneous qPCR reactions may restrict amplification optimisation, introducing bias<br>- PCR inhibitors may impose bias | - Cluster amplification may introduce bias<br>- Bioinformatic workflows can introduce bias, e.g., choice of database |
| **Coverage** | - ARG coverage limited to the number of genes targeted | + Nontargeted, provides a comprehensive profile of the resistome |
| **Novel gene discovery** | - Extremely limited capability for novel gene discovery | + Methods like deepARG and fARGene can identify novel ARGs directly from sequence reads<br>+ Existing datasets can be reanalysed for later discovered ARGs |
| **Contextual information** | - Can target specific taxa and other genes of interest (i.e. MGEs) but overall provides little contextual information | + Provides taxonomic data and can identify potential host bacteria through MAG analysis<br>+ Identification of neighbouring genes can provide contextual information, e.g. mobility potential |
| **Retrospective analysis** | - Incapable | + Datasets can be reanalysed with updated tools and databases |

dynamics can become more apparent (Joseph, Battaglia et al. 2019, Sims, Kannan et al. 2023). Additionally, while the current study only compared ARG relative abundances between methods, quantifying absolute abundances is important for assessing true temporal and spatial ARG variations. Absolute abundances can be derived from the 16S rRNA gene copy number, though this can only be considered a rough estimate. Quantitative metagenomics is possible with the use of internal standard spike-ins (Crossette, Gumm et al. 2021); however, this introduces bias. HT qPCR analysis can achieve more accurate quantification, especially with the inclusion of standard curves for each ARG, albeit at the expense of reduced throughput. Nevertheless, when considered in conjunction with its greater sensitivity, HT qPCR presents as the more suitable approach for surveillance programs demanding precise quantification. This might include those aiming to compare ARG abundances between populations or environments. For example, HT qPCR was the choice method to compare the ARGs abundances in wastewater against clinical levels (Parnanen, Narciso-da-Rocha et al. 2019). Regarding clinical wastewater, both methods found the hospital effluent to harbour a substantially richer and more abundant resistome than that of the WWTP influent. Although this is unsurprising, existing literature provides no strong consensus on the impact of hospital wastewater as a source of ARGs in WWTP influent (Buelow, Bayjanov et al. 2018, Lepper, Perry et al. 2023, Sims, Kannan et al. 2023), supporting the need for AMR monitoring programs involving both hospitals and WWTPs.

Concerning the spatial dynamics of wastewater influent resistome, both approaches found a modest effect of geographic region on composition, with it being most distinct between North and South Wales. Both methods attributed a greater percentage of compositional variation to environmental variables compared to anthropogenic variables, with ammonium and phosphorus concentrations issuing the strongest influence. Acting as nutritional sources, this

may suggest an indirect impact through their influence on the microbial community (Guo, Ni et al. 2017). There was otherwise some discrepancy between methods as to which variables explained a greater degree of variance. Therefore, despite having revealed similar spatiotemporal dynamics in the resistome composition, differences in ARGs relative abundances did affect the interpretation of influencing factors. These results contrast the conclusions drawn by Liu, Xiao (2019), who suggested the impact of the two approaches was negligible when profiling reservoir water. However, they only found one of the tested variables, pH, to correlate with the ARG profile. Determining the most suitable method for investigating potential selective agents depends on the insights desired from the data. Correlating variables against ARG abundances would benefit from the sensitive quantification achieved by HT qPCR, whereas metagenomics could provide genetic context for underlying mechanisms, as discussed below.

Given human pathogens present a greater risk than commensal or environmental bacteria, ARG host identification should be considered a priority of AMR surveillance. Though HT qPCR can quantify targeted taxa, metagenomics offers comprehensive microbial community data, as used in this study. Correlation analysis linked tetracycline resistance genes to the highest number bacteria genera, found concurrently by both approaches. For example, *tetW* and *tetO* correlated with the Firmicutes genera, *Blautia* and *Faecalibacterium*, identified as potential hosts in past studies (Li, Yang et al. 2015, Liu, L., Zhang et al. 2022). The metagenomic data network also revealed relationships between bacteria genera and ARGs outside the HT qPCR target set. But otherwise, there was a strong degree of congruence between the results, indicating that the choice of method would not greatly affect the prediction of potential host bacteria by correlation analysis. However, it is important to note that such correlations cannot be considered causative without further validation. This is only feasible when working with metagenomic data, as it can produce composite contig

assemblies or metagenome-assembled genomes (MAGs). Through their co-location to phylogenetic markers, ARGs can be linked to specific taxa (Zhang, Z., Zhang et al. 2022). Though it must be noted that as composite assemblies, they are not reflective of a true singular microbe, and many ARGs carried on MGEs cannot be assigned to a host. Nonetheless, the assembly of metagenomic data can be applied towards associating ARGs with host bacteria, MGEs and co-located resistance genes. As such, metagenomic data is favourable for delivering on objectives such as human health risk evaluation and evolutionary analysis.

Beyond the technical differences explored here, the practical requirements must also be considered when deciding on an appropriate monitoring strategy. In terms of the expertise required, the increasing availability of commercial kits and companies offering services in metagenomics and HT qPCR has made both methods more accessible to researchers. Concerning bioinformatic analysis, HT qPCR data benefits from requiring lower computational power and bioinformatic expertise, and quicker turnaround times. However, the advancements in user-friendly pipelines are lowering the barrier for metagenomic analysis. HT qPCR currently costs less per sample, as explored by Liu, X., Xiao et al. (2019). That said, metagenomics might be considered more cost effective given the amount of data generated, and as advancements in sequencing technologies continue, the costs are likely to improve.



## 5 Conclusion

This study is the first to compare the use of HT qPCR and metagenomics for AMR monitoring in wastewater at a national scale. It provides valuable guidance for researchers seeking to design effective AMR wastewater surveillance strategies. Metagenomics permitted a comprehensive overview of the resistome and has the potential to provide critical

contextual information for ARG risk evaluation. Together with exhibiting greater sensitivity, HT qPCR was shown capable of reflecting the compositional spatiotemporal dynamics of the full ARG profile. Nonetheless, differences in sensitivity and coverage between methodologies did impact the ARGs detected and their relative abundances, affecting how the data was interpreted regarding influencing factors. The insights gained into the capabilities and resource requirements of each method will help fashion surveillance efforts towards specific goals, strengthening the impact of future AMR monitoring initiatives.

## 6 Acknowledgements

This research was supported by Welsh Government under the Welsh Wastewater Programme (C035/2021/2022). We thank Tony Harrington and Ian Trick at Dŵr Cymru Welsh Water, UK and Elizabeth Franks at Hafren Dyfrdwy for their assistance in setting up wastewater sample collections. We thank and the Welsh Wastewater Monitoring Group, Bangor, UK and the WEWASH sample collection and laboratory team at Cardiff University for their assistance in sample processing, and. members of the genomics research hub and Wales gene park for help with metagenomic sequencing. We thank Steve Cobley at Welsh Government for project supervision and management. We thank Dr Mandy Wootten at the Specialist Antimicrobial Chemotherapy Unit, Public Health Wales, University Hospital of Wales and Prof David Graham, Newcastle University for advice on the AMR gene targets.

## 7 CrediT authorship contribution statement

We describe author contributions to the paper using the CRedit taxonomy. Conceptualisation: MEK, GW, WBP, WM, GC, PK, AJW and DLJ; Data Curation: MEK, GW, WBP, AB, LR, DAP and PK; Formal analysis: MEK, GW, WBP, AB, DAP and PK; Funding Acquisition: GC, ID, PK, KF, DLJ and AJW; Investigation: GW, LR and WM; Methodology: MEK, GW,

WBP, LR, WM, PK, and KF; Project Administration: GW, GC, IB, PK, KF, AJW and DLJ; Resources: ID, PK, AJW and DLJ; Software: MEK, WBP, AB, DAP and PK; Supervision: ID, PK, AJW and DLJ; Validation: MEK, GW, WBP, WM, PK, AJW and DLJ; Visualisation: MEK and WBP; Writing-Original Draft: MEK; Writing-Review & Editing: MEK, GW, WBP, AJW and DLJ. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Data availability**

Data will be made available on request.

## 8 References

BASTIAN, M., HEYMANN, S. and JACOMY, M., 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks, *International AAAI Conference on Weblogs and Social Media* 2009.

BEHLING, A.H., WILSON, B.C., HO, D., VIRTA, M., O'SULLIVAN, J.M. and VATANEN, T., 2023. Addressing antibiotic resistance: computational answers to a biological problem? *Current opinion in microbiology*, **74**, pp. 102305.

BONIN, N., DOSTER, E., WORLEY, H., PINNELL, L.J., BRAVO, J.E., FERM, P., MARINI, S., PROSPERI, M., NOYES, N., MORLEY, P.S. and BOUCHER, C., 2023. MEGARes and AMR++, v3.0: an updated comprehensive database of antimicrobial resistance determinants and an improved software pipeline for classification using high-throughput sequencing.

BRINCH, C., LEEKITCHAROENPHON, P., DUARTE, A.S.R., SVENDSEN, C.A., JENSEN, J.D. and AARESTRUP, F.M., 2020. Long-Term Temporal Stability of the Resistome in Sewage from Copenhagen. *mSystems*, **5**(5), pp. e00841-20.

BUELOW, E., BAYJANOV, J.R., MAJOOR, E., WILLEMS, R.J., BONTEN, M.J., SCHMITT, H. and VAN SCHAIK, W., 2018. Limited influence of hospital wastewater on the microbiome and resistome of wastewater in a community sewerage system. *FEMS microbiology ecology*, **94**(7), pp. 10.1093/femsec/fiy087.

CROSSETTE, E., GUMM, J., LANGENFELD, K., RASKIN, L., DUHAIME, M. and WIGGINTON, K., 2021. Metagenomic Quantification of Genes with Internal Standards. *mBio*, **12**(1), pp. e03173-20.

EWELS, P., MAGNUSSON, M., LUNDIN, S. and KALLER, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics (Oxford, England)*, **32**(19), pp. 3047-3048.

FERREIRA, C., OTANI, S., AARESTRUP, F.M. and MANAIA, C.M., 2023. Quantitative PCR versus metagenomics for monitoring antibiotic resistance genes: balancing high sensitivity and broad coverage. *FEMS Microbes*, **4**, pp. xtad008.

GUO, J., NI, B., HAN, X., CHEN, X., BOND, P., PENG, Y. and YUAN, Z., 2017. Unraveling microbial structure and diversity of activated sludge in a full-scale simultaneous nitrogen and phosphorus removal plant using metagenomic sequencing. *Enzyme and microbial technology*, **102**, pp. 16-25.

HUIJBERS, P.M.C., FLACH, C. and LARSSON, D.G.J., 2019. A conceptual framework for the environmental surveillance of antibiotics and antibiotic resistance. *Environment international*, **130**, pp. 104880.

JOSEPH, S.M., BATTAGLIA, T., MARITZ, J.M., CARLTON, J.M. and BLASER, M.J., 2019. Longitudinal Comparison of Bacterial Diversity and Antibiotic Resistance Genes in New York City Sewage. *mSystems*, **4**(4), pp. e00327-19.

KARKMAN, A., JOHNSON, T.A., LYRA, C., STEDTFELD, R.D., TAMMINEN, M., TIEDJE, J.M. and VIRTA, M., 2016. High-throughput quantification of antibiotic resistance genes from an urban wastewater treatment plant. *FEMS microbiology ecology*, **92**(3), pp. fiw014. doi: 10.1093/femsec/fiw014. Epub 2016 Jan 31.

LAI, F.Y., MUZIASARI, W., VIRTA, M., WIBERG, K. and AHRENS, L., 2021a. Profiles of environmental antibiotic resistomes in the urban aquatic recipients of Sweden using high-throughput quantitative PCR analysis. *Environmental pollution (Barking, Essex : 1987)*, **287**, pp. 117651.

LAI, F.Y., MUZIASARI, W., VIRTA, M., WIBERG, K. and AHRENS, L., 2021b. Profiles of environmental antibiotic resistomes in the urban aquatic recipients of Sweden using high-throughput quantitative PCR analysis. *Environmental pollution (Barking, Essex : 1987)*, **287**, pp. 117651.

LEPPER, H.C., PERRY, M.R., WEE, B.A., WILLS, D., NIELSEN, H., OTANI, S., SIMON, M., AARESTRUP, F.M., WOOLHOUSE, M.E.J. and VAN BUNNIK, B.A.D., 2023. Distinctive hospital and community resistomes in Scottish urban wastewater: Metagenomics of a paired wastewater sampling design. *The Science of the total environment*, **902**, pp. 165978.

LI, B., YANG, Y., MA, L., JU, F., GUO, F., TIEDJE, J.M. and ZHANG, T., 2015. Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. *The ISME journal*, **9**(11), pp. 2490-2502.

LIU, L., ZHANG, Y., CHEN, H. and TENG, Y., 2022. Fate of resistome components and characteristics of microbial communities in constructed wetlands and their receiving river. *The Science of the total environment*, **844**, pp. 157226.

LIU, X., XIAO, P., GUO, Y.Y., LIU, L.M. and YANG, J., 2019. The impacts of different high-throughput profiling approaches on the understanding of bacterial antibiotic resistance genes in a freshwater reservoir. **693**.

MCEWEN, S.A. and COLLIGNON, P.J., 2018. Antimicrobial Resistance: a One Health Perspective. *Microbiology spectrum*, **6**(2), pp. 10.1128/microbiolspec.ARBA-2017.

MUNK, P., BRINCH, C., MOLLER, F.D., PETERSEN, T.N., HENDRIKSEN, R.S., SEYFARTH, A.M., KJELDGAARD, J.S., SVENDSEN, C.A., VAN BUNNIK, B., BERGLUND, F., GLOBAL SEWAGE SURVEILLANCE CONSORTIUM, LARSSON, D.G.J., KOOPMANS, M., WOOLHOUSE, M. and AARESTRUP, F.M., 2022. Genomic analysis of sewage from 101 countries reveals global landscape of antimicrobial resistance. *Nature communications*, **13**(1), pp. 7251-7.

MUZIASARI, W.I., PITKANEN, L.K., SORUM, H., STEDTFELD, R.D., TIEDJE, J.M. and VIRTA, M., 2017. Corrigendum: The Resistome of Farmed Fish Feces Contributes to the Enrichment of Antibiotic Resistance Genes in Sediments below Baltic Sea Fish Farms. *Frontiers in microbiology*, **8**, pp. 1491.

PARNANEN, K.M.M., NARCISO-DA-ROCHA, C., KNEIS, D., BERENDONK, T.U., 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., SORUM, H., STEDTFELD, R.D., TIEDJE, J.M., GIUSTINA, S.V.D., WALSH, F., VAZ-MOREIRA, I., VIRTA, M. and MANAIA, C.M., 2019. Antibiotic resistance in European wastewater treatment plants mirrors the pattern of clinical antibiotic resistance prevalence. *Science advances*, **5**(3), pp. eaau9124.

PRUDEN, A., VIKESLAND, P.J., DAVIS, B.C. and DE RODA HUSMAN, A.M., 2021. Seizing the moment: now is the time for integrated global surveillance of antimicrobial resistance in wastewater environments. *Current opinion in microbiology*, **64**, pp. 91-99.

R CORE TEAM, 2022-last update, R: A Language and Environment for Statistical Computing [Homepage of R Foundation for Statistical Computing], [Online]. Available: https://www.R-project.org/.

ROBINS, K., LEONARD, A.F.C., FARKAS, K., GRAHAM, D.W., JONES, D.L., KASPRZYK-HORDERN, B., BUNCE, J.T., GRIMSLEY, J.M.S., WADE, M.J., ZEALAND, A.M. and MCINTYRE-NOLAN, S., 2022. Research needs for optimising wastewater-based epidemiology monitoring for public health protection. *Journal of water and health*, **20**(9), pp. 1284-1313.

SIMS, N., KANNAN, A., HOLTON, E., JAGADEESAN, K., MAGEIROS, L., STANDERWICK, R., CRAFT, T., BARDEN, R., FEIL, E.J. and KASPRZYK-HORDERN, B., 2023. Antimicrobials and antimicrobial resistance genes in a one-year city metabolism longitudinal study using wastewater-based epidemiology. *Environmental Pollution*, **333**, pp. 122020.

STEDTFELD, R.D., GUO, X., STEDTFELD, T.M., SHENG, H., WILLIAMS, M.R., HAUSCHILD, K., GUNTURU, S., TIFT, L., WANG, F., HOWE, A., CHAI, B., YIN, D., COLE, J.R., TIEDJE, J.M. and HASHSHAM, S.A., 2018. Primer set 2.0 for highly parallel qPCR array targeting antibiotic resistance genes and mobile genetic elements. *FEMS microbiology ecology*, **94**(9), pp. fiy130. doi: 10.1093/femsec/fiy130.

STEEN, A.D., CRITS-CHRISTOPH, A., CARINI, P., DEANGELIS, K.M., FIERER, N., LLOYD, K.G. and CAMERON THRASH, J., 2019. High proportions of bacteria and archaea across most biomes remain uncultured. *The ISME journal*, **13**(12), pp. 3126-3130.

WEBSTER, G., NEWBERRY, C.J., FRY, J.C. and WEIGHTMAN, A.J., 2003. Assessment of bacterial community structure in the deep sub-seafloor biosphere by 16S rDNA-based techniques: a cautionary tale. *Journal of microbiological methods*, **55**(1), pp. 155-164.

WEI, Z., FENG, K., LI, S., ZHANG, Y., CHEN, H., YIN, H., XU, M. and DENG, Y., 2018. Exploring abundance, diversity and variation of a widespread antibiotic resistance gene in wastewater treatment plants. *Environment international*, **117**, pp. 186-195.

WILDE, H., PERRY, W.B., JONES, O., KILLE, P., WEIGHTMAN, A., JONES, D.L., CROSS, G. and DURANCE, I., 2022. Accounting for Dilution of SARS-CoV-2 in Wastewater Samples Using Physico-Chemical Markers. *Water*, **14**(18),.

WOOD, D.E., LU, J. and LANGMEAD, B., 2019. Improved metagenomic analysis with Kraken 2. *Genome biology*, **20**(1), pp. 257-0.

ZHANG, A., GASTON, J.M., DAI, C.L., ZHAO, S., POYET, M., GROUSSIN, M., YIN, X., LI, L., VAN LOOSDRECHT, M.C.M., TOPP, E., GILLINGS, M.R., HANAGE, W.P., TIEDJE, J.M., MONIZ, K., ALM, E.J. and ZHANG, T., 2021. An omics-based framework for assessing the health risk of antimicrobial resistance genes. *Nature communications*, **12**(1), pp. 4765-3.

ZHANG, Z., ZHANG, Q., WANG, T., XU, N., LU, T., HONG, W., PENUELAS, J., GILLINGS, M., WANG, M., GAO, W. and QIAN, H., 2022. Assessment of global health risk of antibiotic resistance genes. *Nature communications*, **13**(1), pp. 1553-8.

ZHENG, W., HUYAN, J., TIAN, Z., ZHANG, Y. and WEN, X., 2020. Clinical class 1 integron-integrase gene – A promising indicator to monitor the abundance and elimination of antibiotic resistance genes in an urban wastewater treatment plant. *Environment international*, **135**, pp. 105372.