Journal of Microbiological Methods 228 (2025) 107069

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Concordance in molecular methods for detection of antimicrobial resistance: A cross sectional study of the influent to a wastewater plant

Kezia Drane [a], Roger Huerlimann [b,c], Rhondda Jones [d], Anna Whelan [e], Madoc Sheehan [c], Ellen Ariel [a], Robert Kinobe [a,*]

[a] College of Public Health, Medical and Veterinary Sciences, 1 Solander Drive, James Cook University, Townsville, Queensland 4811, Australia
[b] Marine Climate Change Unit, Okinawa Institute of Science and Technology, 1919-1 Tacha, Onna-son, Okinawa 904-0495, Japan
[c] College of Science and Engineering, 1 James Cook Drive, James Cook University, Townsville, Queensland 4811, Australia
[d] Graduate Research School, 1 James Cook Drive, James Cook University, Townsville, Queensland 4811, Australia
[e] Townsville Water and Waste, Wastewater Operations, Townsville, Queensland 4810, Australia

ARTICLE INFO

Keywords:
Wastewater
Antimicrobial resistance
Next-generation sequencing
RNA sequencing
Microbiome
Antibiotics

ABSTRACT

Methods that are used to characterise microbiomes and antimicrobial resistance genes (ARGs) in wastewater are not standardised. We used shotgun metagenomic sequencing (SM-Seq), RNA sequencing (RNA-seq) and targeted qPCR to compare microbial and ARG diversity in the influent to a municipal wastewater treatment plant in Australia. ARGs were annotated with CARD-RGI and MEGARes databases, and bacterial diversity was characterised by 16S rRNA gene sequencing and SM-Seq, with species annotation in SILVA/GreenGenes databases or Kraken2 and the NCBI nucleotide database respectively. CARD and MEGARes identified evenly distributed ARG profiles but MEGARes detected a richer array of ARGs (richness = 475 vs 320). Qualitatively, ARGs encoding for aminoglycoside, macrolide-lincosamide-streptogramin and multidrug resistance were the most abundant in all examined databases. RNA-seq detected only 32 % of ARGs identified by SM-Seq, but there was concordance in the qualitative identification of aminoglycoside, macrolide-lincosamide, phenicol, sulfonamide and multidrug resistance by SM-Seq and RNA-seq. qPCR confirmed the detection of some ARGs, including *OXA*, *VEB* and *EREB*, that were identified by SM-Seq and RNA-seq in the influent. For bacteria, SM-Seq or 16S rRNA gene sequencing were equally effective in population profiling at phyla or class level. However, SM-Seq identified a significantly higher species richness (richness = 15,000 vs 3750). These results demonstrate that SM-Seq with gene annotation in CARD and MEGARes are equally sufficient for surveillance of antimicrobial resistance in wastewater. For more precise ARG identification and quantification however, MEGARes presented a better resolution. The functionality of detected ARGs was not confirmed, but general agreement on the putative phenotypic resistance profile by antimicrobial class was observed between RNA-Seq and SM-Seq.

1. Introduction

The dissemination of antibiotics and antibiotic resistance genes (ARGs) into aquatic ecosystems is a major global concern, as it promotes the selection of antibiotic resistant bacteria (ARB) within complex microbial populations (Guo et al., 2017; Tiwari et al., 2022; Bonetta et al., 2023). Wastewater treatment plants (WWTPs) are a well characterised point of convergence of antibiotics, bacteria and ARGs, and their subsequent dissemination into aquatic ecosystems (Adekanmbi et al., 2020; Hendriksen et al., 2019; Liu et al., 2019a; Muurinen et al., 2017; Wang et al., 2014). Currently, many different molecular techniques are used to quantify as well as describe qualitative features of bacteria and ARGs in WWTPs, but there is no standardised approach and cross-validation of methods is rarely done (Gholipour et al., 2024). This makes it difficult to qualitatively and quantitatively compare, or validate results from different studies. Customised primers in polymerase chain reactions (PCR) or quantitative PCR (qPCR) are commonly used to to detect and quantify specific ARGs at low copy numbers thus providing very high sensitivity (Rocha et al., 2019; Liu et al., 2019a). However, PCR and qPCR assays can take time to optimise and the scope of ARG detection is usually limited by experimental design or primer sets provided on commercial arrays (Gaviria-Figueroa et al., 2019; Muurinen et al.,

* Corresponding author.
E-mail address: Robert.kinobe@jcu.edu.au (R. Kinobe).

https://doi.org/10.1016/j.mimet.2024.107069
Received 28 December 2023; Received in revised form 1 October 2024; Accepted 15 November 2024
Available online 17 November 2024

2017).

By contrast, sequencing of the 16S ribosomal ribonucleic acid (16S rRNA) gene and, shotgun metagenomic DNA sequencing (SM-Seq) approaches have been used to identify entire microbiomes or a much wider array of ARGs in experimental samples (Christgen et al., 2015; Guo et al., 2017; Majeed et al., 2021). SM-Seq or RNA sequencing and probing of known DNA or RNA libraries of ARG databases is the method of choice for the global surveillance of markers for antimicrobial resistance in sewage and other environmental matrices (Hendriksen et al., 2019; Larsson and Flach, 2022). SM-Seq is now readily available since the cost has declined through economy of scale and it is a robust and efficient method because all major steps of the process can be automated. SM-Seq can also be used to retrospectively analyse archived samples to identify novel ARGs or ARG mutations (Boolchandani et al., 2019; Christgen et al., 2015; Guo et al., 2017; Majeed et al., 2021). Nonetheless, some limitations to the applicability of SM-Seq for ARG detection still exist. Highly technical knowledge and expertise is required for several steps involved in SM-Seq approaches, and detection of ARGs using this method is limited by the scope of reference ARG sequences in probed databases such as the Comprehensive Antibiotic Resistance Database (CARD), Antibiotic Resistance Gene-ANNOTation (ARG-ANNOT), an antimicrobial resistance database for high throughput sequencing (MEGARes), and an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data (ResFinder) (Hendriksen et al., 2019). In addition, neither DNA sequencing nor PCR/qPCR provides a direct link to the functional nature of identified genetic determinants unless it is coupled with other forms of identification such as culture and sensitivity testing or transcriptomic analyses (Boolchandani et al., 2019; Gaviria-Figueroa et al., 2019). For example, in a comparative study targeting specific ARGs against defined antimicrobial classes in faecal coliforms, qPCR data strongly correlated with culture based evaluations in wastewater, recycled water and tap water (Rocha et al., 2019). Likewise, high concordance was observed between three different SM-Seq approaches for ARG or virulence detection, and phenotypic expressions in culture-based methods (Mason et al., 2018). However, data validation based on bacterial culture is only effective for specific target ARGs and bacterial isolates. Its application to validate data for samples from complex environmental matrices, with a diverse bacterial population, and a much bigger ARG pool is not practical. As such, robust molecular methods including qPCR, SM-Seq and RNA sequencing are more likely to be used for high throughput screening of complex environmental samples, but direct information on agreement in different molecular methods is scarce. The primary objective of this study was to compare qPCR, SM-Seq and RNA sequencing in the detection and relative quatification of ARGs, and the associated bacterial population in the influent to a metropolitan WWTP. Influent samples are ideal for these comparisons because of the likely accurate representation of microbial and ARG diversity, free of wastewater processing influences such as biological treatment reactions and chemical additions.

2. Materials and methods

2.1. Study site and sampling

Samples of influent were obtained from the Cleveland Bay Sewage Treatment plant (CBSTP), Queensland, Australia, with the help of the Townsville City Council (TCC). The city of Townsville lies on the eastern coast of Australia adjacent to the boundary of the Great Barrier Reef Marine Park (GBRMP). The CBSTP is a membrane bioreactor plant that discharges processed effluent directly into the Queensland State Marine Park and within the general use zone of the GBRMP located at −19.288744118517624 and 146.8551011782239. The CBSTP is Townsville's largest municipal sewage treatment plant, and currently services up to approximately 110,000 equivalent persons and it utilises the largest membrane bioreactor in the southern hemisphere. The catchment for the influent into the CBSTP includes commercial, industrial, and domestic sources as well as a multitude of medical and veterinary sources all of which may influence the ARG and ARB load. Samples of CBSTP influent were collected daily over 18 consecutive days and then used to create a composite sample that captures any short-term temporal fluctuation in the influent. Samples were collected in sterile one litre containers using an autosampler and then transported to the laboratory within 1 h at 5 °C for further processing.

2.2. Nucleic acid extraction and preservation

Samples (1–2 mL aliquots) were centrifuged at 10,000 ×*g* for 2 min at 25 °C. The supernatant was then discarded, and the pellet was preserved with 750 μL of DNA /RNA shield. The pelleted 2 mL aliquots were pooled into two composite samples and DNA or RNA was extracted using ZymoBIOMICS™ Miniprep kits (Zymo Research Corp., Irvine, CA, USA) according to the manufacturer's instructions. The concentration and purity of DNA or RNA was determined using a NadoDrop™ ND-2000c spectrophotometer (Thermo Fisher Scientific Inc., MA USA). Extracted nucleic acids were stored at −20 °C (DNA) or − 80 °C (RNA) until further analyses. As environmental based nucleic acids are notoriously difficult to extract at workable quantities, extracted DNA and RNA was considered suitable for downstream applications at a concentration ≥ 10 ng/μL. All samples had acceptable quality for DNA and RNA; theses were a 260/280 nm ratio = 1.80 ± 0.2 and 2.00 ± 0.2 respectively, and a 260/230 ratio ≥ 2 for both DNA and RNA. Extracted DNA was used for qPCR, 16S rRNA gene sequencing, and SM-Seq, while extracted RNA was subjected to RNA sequencing for transcriptomic analysis.

2.3. Using qPCR to detect and quantify ARGs in pooled influent samples for a WWTP

Extracted DNA was pooled into two composite samples for these analyses. Microbial DNA qPCR analysis for ARGs was done using customised arrays (cat. No. 330261 BAID-1901ZRA, Qiagen, Valencia, CA USA). These arrays provide a high-throughput profiling of 83 genes that represent major classes of ARGs including aminoglycosides, beta-lactams, fluoroquinolones, macrolides-lincosamides-streptogramins, tetracyclines, vancomycin, and other drug classes (Supplementary Table 1). Thermal cycling was performed on an applied Biosystems™ QuantStudio™ 3 Real-Time PCR System thermocycler (Waltham, MA, USA). Five nanograms of template DNA was used with a Hot Start Taq DNA polymerase and an initial PCR activation at 95 °C for 10 mins followed by 40 cycles consisting of denaturation for 15 s at 95 °C and annealing and extension for 2 min at 60 °C. A threshold cycle (C_T) value of 37 was used and baselines were manually set for cycles 8–20 with a threshold fluorescence setting of 0.24 and 0.35 respectively for each duplicate. qPCR array data was evaluated according to the manufacturer's criteria and as previously published (Adekanmbi et al., 2020; Liu et al., 2018; Liu et al., 2019a; Looft et al., 2012; Muurinen et al., 2017).

2.4. Shotgun metagenomic DNA or RNA sequencing to determine bacterial population and ARG profiles

DNA extracted from 18 samples based on 18 days of sampling was pooled into a composite sample and then subjected to SM-Seq at the Australian Genome Research Facility (AGRF), Melbourne Australia. Library preparation and SM-Seq was performed according to AGRF's methodology. Briefly, the Shotgun library was constructed by using Illumina Nextera XT DNA library preparation kit (San Diego, CA, USA). DNA was fragmented to a size of 150 bp, and the fragments were then ligated with index barcode sequences, end-polished, A-tailed and ligated with adaptors for Illumina sequencing. The library was sequenced using Illumina NovaSeq S4 sequencing platform (San Diego, CA, USA) with a single lane for this sample. The raw read sequences were processed through Trim Galore (version 0.6.5) and cutadapt (V2.10) to identify

and remove sequencing adapters and low-quality stretches of base quality below 30 as outlined previously (Kruerger, 2015; Martin, 2011). The Kraken2-built script tool in Kraken2 package (version 2.0.8) (Wood et al., 2019), and Bracken (version 2.5) (Lu et al., 2017) were applied to profile the composition of microbial communities using NCBI genome and nucleotide (nt) database (downloaded 15/04/2021). Gene fraction was defined as the minimum proportion of nucleotides in a reference sequence that had to be aligned by at least one read to be considered as 'identified' in the sequenced data. A threshold of 80 % was used to minimize potential false classifications and removing accessions with sparse alignment and functional profiles were generated using HUMAnN2 (version 2.8.1).

ARGs were identified from SM-Seq data using two separate analysis pipelines as outlined in Fig. 1. AMR++ 3.0 was used for quality assessment of raw reads, followed by filtering, trimming and adapter removal as previously described (Bonin et al., 2023). This pipeline also facilitated the removal of human DNA. Two separate databases were then probed to identify ARGs. Using AMR++ 3.0 facilitated the probing of MEGARes database using bwa-mem as described previously (Bonin et al., 2023). ResistomeAnalyzer was then applied to process the alignments using a minimum gene fraction threshold of 80 %. Further ARG analysis was also completed using the CARD resistance gene identifier (RGI) (version 6.0.2) (Alcock et al., 2023). CARD-RGI aligned short DNA sequences in FASTQ format using KMA (version 1.3.4) and then the FASTQ sequences were aligned to curated reference sequences and in silico predicted allelic variants in the CARD database (version3.2.6) (Clausen et al., 2018). This alignment of the CARD database sequences was done using the BWT function and both strict and perfect alignment paradigms in CARDI RGI. For the aligned read data, gene fraction threshold was set at 80 % to minimize potential false classifications. Finally, a resistome matrix was built using the two databases to get final results. These two databases were selected based on their comprehensiveness of ARG coverage, frequency of maintenance and update, and being publically accessible. MEGARes contains approximately 9000 ARGs and the CARD database contains 5010 ARG reference sequences, and this is updated regularly (Alcock et al., 2023; Bonin et al., 2023).

RNA was extracted from a composite sample comprising of 18 days of sampled influent, corresponding to the same 18 days used for DNA analysis described above. The extracted RNA composite sample was then subjected to RNA sequencing at the AGRF. Library preparation and sequencing was performed according to AGRF's methodology. Briefly, a library was created using Illumina Stranded Total RNA with Ribo-Zero Plus kit (San Diego, CA, USA) using 150 ng input, 16 PCR cycles and a 0.8× bead ratio for the final library clean-up.

The quality of the library was assessed using the LabChip GX Touh II DNA High Sensitivity Assay (Perkin Elmer, Hamburg, Germany), and the concentration of the library was determined using NEB's qPCR kit and sequenced on Illumina's NovaSeq 6000 (San Diego, CA, USA) on a lane of an S4 300 flow cell. Data was processed through AMR++3.0 for identification of functional ARGs in the sample as described previously (Bonin et al., 2023).

2.5. 16S rRNA gene sequencing to determine the bacterial population

16S rRNA gene sequencing was performed according to AGRF's methodology. Briefly, the V3-V4 hypervariable region of the 16S rRNA gene was amplified using universal primers 341-Forward (5'CCTACGGGNGGCWGCAG) and 806-Reverse (5' GGACTACHVGGGTWTCTAAT). DNA quality control screening was done via PCR and indexed using fluorometry prior to diversity profiling of the 16S rRNA primer targets. Amplicons were sequenced on the Illumina MiSeq platform (San Diego, CA, USA) for the production of a 300 bp paired end run. Diversity profile analysis was performed with QIIME 22019.7 (Bolyen et al., 2019). The demultiplexed raw reads were primer trimmed and quality filtered using the cutadapt plugin followed by denoising and removal of chimeras via q2-dada2 as previously described (Callahan et al., 2016). Taxonomy was assigned from the SILVA database (version 138.1) (Quast et al., 2012) and GreenGenes database (version 13.5) (Desantis et al., 2006) to amplicon sequence variants using the q2-feature-classifier (Bokulich et al., 2018), and sequence variant set at 97 % identity.

2.6. Data analysis

2.6.1. Analysis of ARG expression by qPCR

The criteria for qPCR data analysis were: (1) the sample was identified as positive for a gene if the C_T was below 34, (2) detection for a gene was inconclusive if the C_T value was between 34 and 37, and (3)

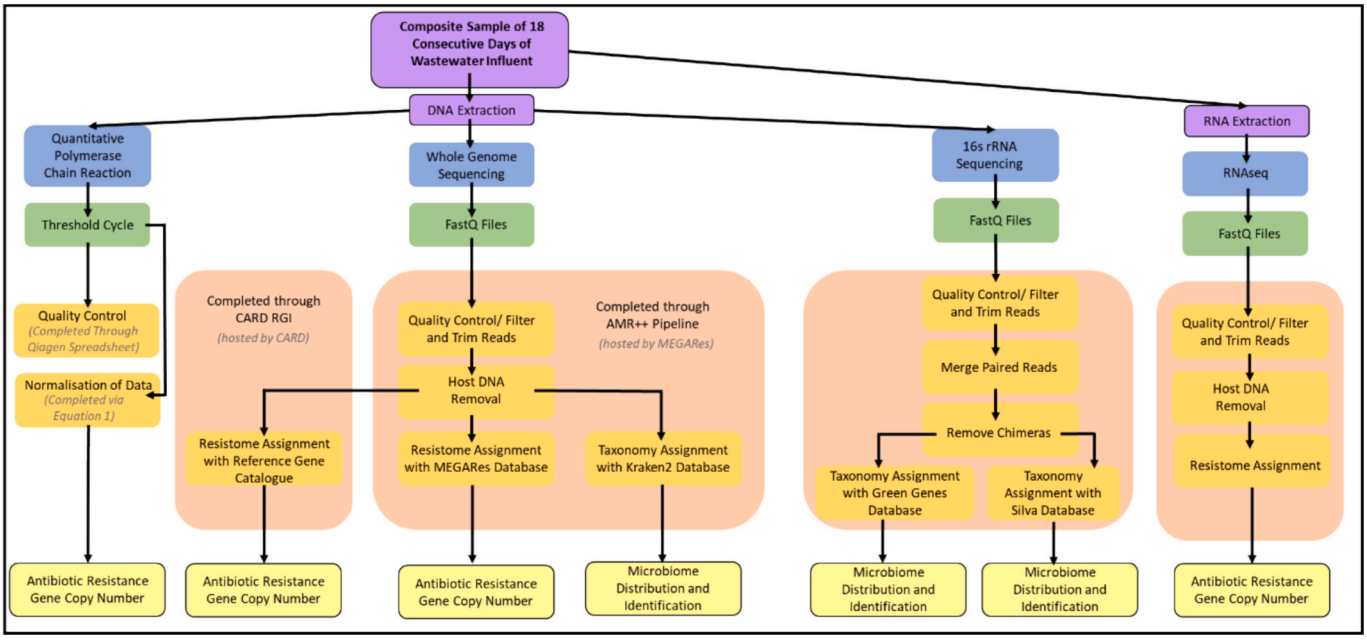

Fig. 1. A simplified overview of different methods and analyses that were used. This outlines the workflow from sampling to the final qualitative or quantitative results including antibiotic resistance gene copy numbers and microbiome identification and distribution.

the sample was identified as negative for a gene if the C_T value was greater than 37. These guidelines were based on the C_T values obtained with the negative control sample prepared with molecular grade water. The C_T value for each positively identified gene was converted to gene copy number using Eq. (1) as previously described (Adekanmbi et al., 2020; Liu et al., 2018; Liu et al., 2019a; Looft et al., 2012; Muurinen et al., 2017).

$$\text{Relative Gene Copy} = 10^{(34-C_T)\div\left(\frac{10}{3}\right)} \quad (1)$$

Where C_T refers to number of cyles required to exceed background fluorescence for each probed gene, 34 refers to the detection limit, and 10/3 refers to the 10-fold difference in gene copy numbers (GCN) at 100 % efficiency.

2.6.2. Analysis of ARG expression by shotgun metagenomic DNA and RNA sequencing

Cumulative read counts of each ARG were normalized to gene length as reads per kilobase (RPK) as outlined previously (Honda et al., 2023). The relative proportion of each ARG in the sample was evaluated as the ratio of RPK for each ARG to the sum of RPK for all ARGs. Data were processed and presented using the "ggplot2" package version 3.4.1 (Wickham, 2011). Statistical analysis of diversity and richness of ARG expression was assessed using the "Phyloseq" package version 3.16 (Mcmurdie and Holmes, 2012). Alpha diversity of ARGs included the evaluation of Richness and Eveness. Differences in detection methods for relative ARG expression in the composite sample were also analysed by a correlation coefficient model (R software version 4.1.1) and *Z*-scores.

2.6.3. Analysis of bacterial population profiles

The microbiome composition identified by probing the NCBI nucleotide database using SM-Seq Kraken2 software, and two 16S rRNA sequence databases (GreenGenes and SILVA) were also compared using R software (v4.1.1) as previously described (Odom et al., 2023). Read counts of 16S rRNA gene were also normalized as RPK and the relative proportions of different classes of bacteria was evaluated as the ratio of RPK for each class to the sum of RPK for the sample and presented as bar plots using "ggplot2" package version 3.4.1 (Wickham, 2011). "Phyloseq" package (version 3.16) was used for alpha diversity assessments as outlined above.

3. Results

3.1. Characterisation of the ARG profile by a customised microbial DNA qPCR array

Bacterial phosphoenolpyruvate carboxylase gene (*PPC*) was used as a positive control for each qPCR array and it was amplified within the recommended C_T of 22 ± 2 for the duplicate assays. Duplicate DNA qPCR arrays identified a similar ARG profile in the composite sample. The averages of C_T values from duplicate arrays showed that 45 out of 83 individual ARGs and variants were positively expressed while the expression of 10 genes was categorised as inconclusive and 28 gene were not expressed (Supplementary Table 1). For the 45 ARGs and variants expressed in the composite influent sample, there was no difference in duplicate C_T values from two separate assays. The average difference in C_T values was 0.42 with a range of 0.04 (for *CCRA* gene) to 1.17 (for *QNRB-1* gene). Average C_T values were used to determine gene copy number as outlined in (Eq. (1)). Of the 83 ARGs screened in duplicate qPCR assays, 10 were represented multiple times either as specific variants of the same gene e.g., *CTX-1*, *CTX-8* and *CTX-9*, or sequences of the same gene with reference to different bacterial strains of origin e.g., *SHV*(238G240E) and *SHV*(238G240K) (Supplementary Table 1). Copy numbers of variants of the same gene were aggregated and represented as a single gene and the type of resistance it encodes for. The 10 most abundant ARGs were *OXA* (31.0 %), *ERMB* (19.1 %), *GES* (15.2 %), *MEFA* (15.0 %), *VEB* (4.6 %), *TETA* (4.0 %), *AADA1* (3.9 %), *AAC6* (2.6 %), *QNRS* (1.3 %) and *MOX* (0.8 %), collectively making up 97.5 % of all detected ARGs (Fig. 2). Resistance genes associated with beta-lactamases had the highest percentage of GCN (53 %) and this may be attributable to high representation (28 out of 50) ARG primer sets to probe the expression of beta-lactamases. In contrast, the combined GCN for five ARGs encoding for resistance against macrolides, lincosamides and streptogramins (MLS) accounted for 35 % of all detected GCN. Other represented antimicrobial classes were fluoroquinolones, tetracyclines and aminoglycosides.

3.2. ARG characterisation by illumina sequencing and probing of annotated datases

3.2.1. DNA sequencing and ARG annotation in MEGARes database

The MEGARes database identified 184 ARGs out of 487,068 reads and each ARG sequence matched at least 80 % of the individual gene fraction (GF) in the database. ARGs encoding for resistance against the aminoglycosides class of antimicrobials were the most abundant representing 33 %, but other major antimicrobial classes including MLS (31 %), biocides (11 %), beta-lactams (6 %), fluoroquinolone (4 %), and tetracyclines (7 %) were also represented (Fig. 3). To provide context, genes encoding for aminoglycoside resistance represented 19 % of all ARGs in MEGARes database while ARGs against MLS represented less than 5 % of ARGs in the database. A single gene, *A16S* encoding for an aminoglycoside-resistant 16S ribosomal subunit protein accounted for a large proportion (17 %) of the total ARGs in the composite influent sample. Similarly, *MLS23S* encoding for macrolide resistance 23S rRNA mutation accounted for 16 % of the total ARGs in the composite influent sample.

3.2.2. DNA sequencing and ARG annotation in CARD database

The CARD database identified 317 unique ARG sequences from a total of 53,411 reads, and all ARG sequences matched 80–100 % of the individual GF. In this database, genes encoding for multi-drug resistance represented the highest fraction (~40 %) of all ARGs identified in the composite influent sample (Fig. 4). In the distinct antimicrobial classes, resistance against MLS represented the highest fraction (15 %) of all identified ARGs and resistance against other major antimicrobial classes was represented by aminoglycosides (10 %), sulfonamides (7 %), beta-lactams (5 %) and tetracyclines (4 %). Interestingly, this database also identified genes encoding for resistance against disinfectants and antiseptics representing ~11 % of ARGs in CARD and noted herein as drug and biocide resistance (Fig. 4). The *ACRB* gene that encodes for a cell membrane drug antiporter and multi-drug resistance was significantly more abundant representing 32 % of ARG sequences and others included *SUL1* (a sulfonamide resistant dihydropteroate synthase), *SAT4* (streptothricin acetyltransferase) and *QACG* (quaternary ammonium compound gene) representing 6.4 %, 6.0 % and 4 % of total ARGs respectively.

3.2.3. Comparison of qPCR and shotgun metagenomic DNA sequencing for ARG detection

The qPCR array was customised to probe a total of 83 ARGs; out of this, 46 ARGs were in the MEGARes database and 47 ARGs were in the CARD database. Only 3 ARGs (*AAC6*, *AACC* and *APHA1*, encoding for aminoglycoside resistance), were not in the MEGAREs and CARD databases at the time of analysis. Out of all the 46 ARGs that were identifiable by qPCR, MEGARes and CARD, 10 ARGs were detected by qPCR and MEGARes, while 9 ARGs were detected by qPCR and CARD (Fig. 5). Interestingly, the *ERMA* gene that encodes for MLS resistance was detected by MEGARes and CARD at a relatively high abundance but it was not detected by qPCR. The three methods based on SM-Seq and qPCR yielded markedly different GCN, but GCN normalisation using *Z*-scores showed very good agreement in ARG detection and relative

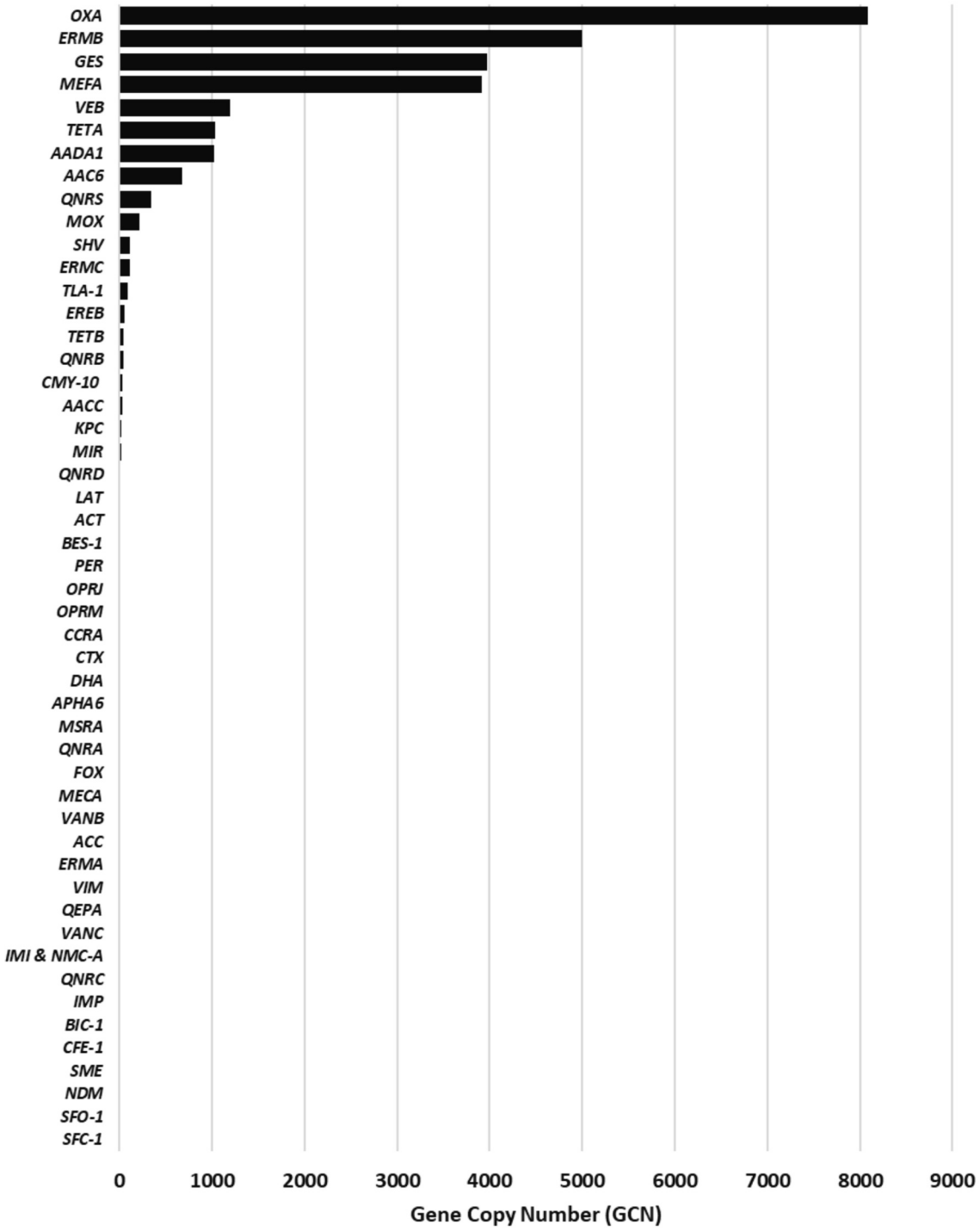

Fig. 2. Average gene copy number of antibiotic resistance genes represented on a qPCR array. DNA was extracted from 18 samples of influent to a wastewater plant. The samples were pooled into a composite sample and assayed in duplicates on two different arrays.

quantification by MEGARes, CARD and qPCR (Fig. 5). For instance, all 3 methods indicate that *OXA*, *GES* and *MEFA* were some of the specific ARGs with the highest relative abundance in respective datasets. Very limited discordance was observed for some ARGs that had very low GCN. The *TETB* and *MOX* genes for example, were detected by qPCR but not via either SM-Seq database (along with an additional 14 ARGs), while *ERMA* and *SME* were detected by one or both SM-Seq methods and not the qPCR array (Fig. 5).

Qualitatively, SM-Seq and gene annotation in MEGARes or CARD identified a similar distribution of ARGs against the major antimicrobial classes albeit at different relative abundance (Fig. 6A). For example, genes encoding for multi-drug resistance represented the highest proportion of ARGs detected in CARD, while genes encoding for aminoglycoside and MLS represented the highest proportion of ARGs detected in MEGARes (Fig. 6A). These observations were also supported by further indepth analysis of ARG distribution in MEGARes and CARD databases (Fig. 6B). Expressed ARGs were more evenly distributed in the CARD database than the MEGARes database (Evenness = 0.32 vs 0.25 respectively). In contrast, MEGARes presented a richer distribution of ARGs than CARD (observed richness = 475 vs 320 ARG sequences

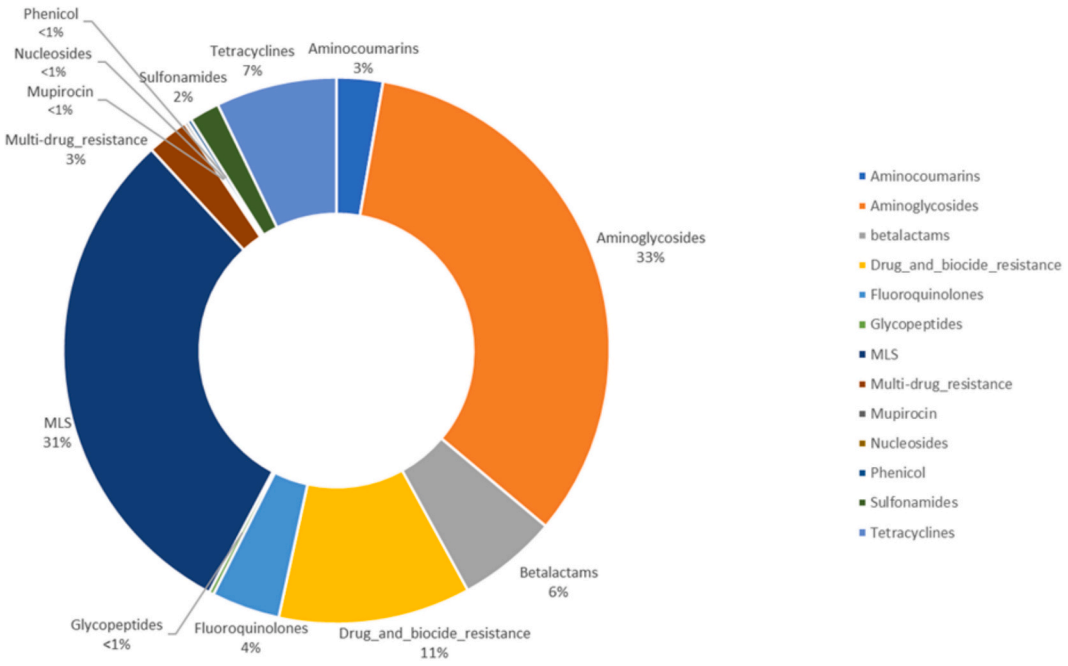

Fig. 3. Distribution of ARGs stratified by antimicrobial class against which they encode resistance after SM-Seq and gene annotation for ARGs in MEGARes databases. This graph denotes ARGs and corresponding antimicrobial classes common to MEGARes and CARD databases for comparison. DNA was extracted from 18 influent samples and then pooled into a composite sample that was analysed by SM-Seq.

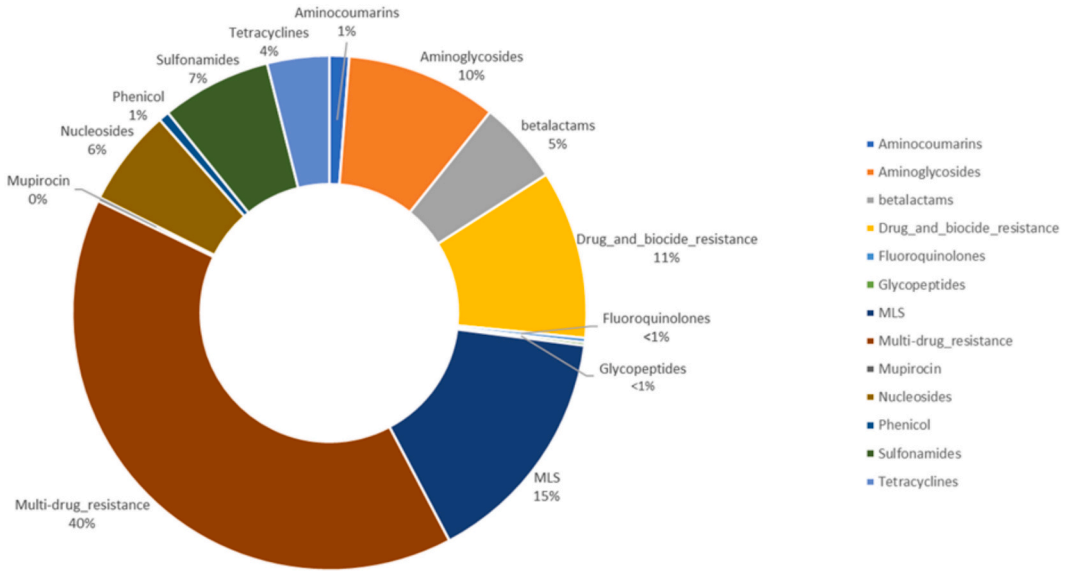

Fig. 4. Distribution of ARGs stratified by antimicrobial class against which they encode resistance after SM-Seq and gene annotation in the CARD database. This graph denotes ARGs and corresponding antimicrobial classes common to MEGARes and CARD databases for comparison. DNA was extracted from 18 influent samples, pooled into a composite sample, and then analysed by SM-Seq.

respectively).

3.2.4. Comparison of RNA sequencing and ARG annotation in MEGARes database

A total of 59 ARGs were identified from RNA sequencing followed by probing of MEGARes in the composite influent sample. This represented 32 % of the total number of ARGs identified by DNA sequencing and gene annotation in MEGARes. All 59 ARGs identified by RNA sequencing were also detected in the DNA sample but the relative proportion of matched reads in RNA for each ARG was generally lower (Fig. 7A). The distribution of all 59 specific ARGs that were identified by sequencing and gene annotation of both DNA and RNA is illustrated in (Fig. 7B). ARGs with the highest abundance in both RNA and DNA sequences were *MLS23S* and *A16S*, encoding for resistance against MLS and aminoglycosides respectively. The magnitude of differences in relative abundance for ARGs expressed in RNA and DNA sequences was illustrated using visualised model coefficients (Fig. 8). These data also show that some uncommon ARGs including *P16S* encoding for pactamycin resistance, and *P23S* encoding for pleuromutilin resistance were represented by a high proportion in RNA sequences as compared to their expression in the DNA sequences. By contrast however, there was no significant difference in the relative abundance between RNA and DNA sequences,

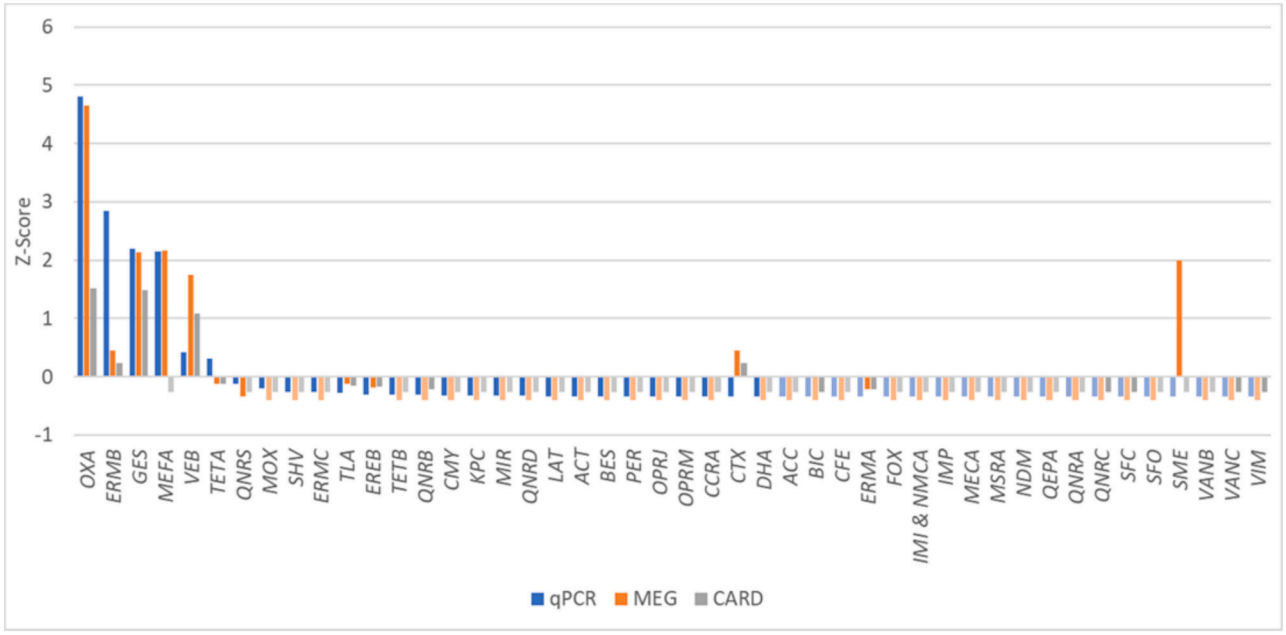

Fig. 5. *Z*-score for normalized gene copy numbers of ARGs identified by qPCR (blue) and SM-Seq followed by gene annotation in the MEGARes (orange) and CARD (grey) databases. Bars represented by lighter, faded colours indicate Z-scores equal to zero actual abundance. DNA was extracted from 18 influent samples, pooled into a composite sample, and then analysed by qPCR and the respective metagenomic DNA sequencing approaches. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

for genes encoding for phenicol, sulfonamide and multi-drug resistance (Figs. 7 and 8).

3.3. Profiling the microbiome by 16S rRNA and shotgun metagenomic sequencing using Kraken2

Profiling of the 16S rRNA amplicon in SILVA database identified 1471 bacterial species while the GreenGenes database identified 1460 bacterial species in the composite influent sample (Fig. 9A). The distribution and relative abundance of bacterial classes was similar for SILVA_16S and GreenGenes_16S methods and this was further illustrated by evaluated alpha diversity for observed population richness and evenness (Fig. 9B). In comparison, a SM-Seq approach coupled with bacterial species profiling in the NCBI nucleotide database using Kraken2 software (SM-Seq_Kraken2) yielded more mapped reads (17,560,817) and bacterial taxonomic sequences (36,255). The relative proportion of some of the most prevalent bacterial classes including Betaproteobacteria, Actinobacteria, Deltaproteobacteria, Fibrobacteria and Flavobacteria was higher in the sample analysed by SM-Seq followed by profiling in Kraken2 (Fig. 9). This difference was also highlighted by profiling bacterial population diversity where SM-Seq_Kraken2 had a markedly higher observed richness of bacterial species (richness = 15,000 for SM-Seq_Kraken2 vs 3750 for 16S rRNA gene metabarcoding). The observed evenness in bacterial species distribution were 0.865 for 16S rRNA gene metabarcoding versus 0.765 for SM-Seq_Kraken2. Further analysis of the 18 individual samples that constituted the composite sample revealed a similar microbial profile as detected in the composite sample using the SILVA_16S rRNA gene metabarcoding approach (Fig. 10). Bacteriodia, Clostridia, Gammaproteobacteria and Synergistia were the bacterial classes with highest abundance across all 18 days of sampling albeit in different proportions. The biggest variation in individual days was in the sample collected on day 15 in which Gammaproteobacteria constituted more than 80 % of the identified species. Analysis of the 18 individual samples as well as the composite sample by a non-parametric, permutational multivariate analysis (PERMANOVA) revealed no significant differences ($P = 0.941$), in the microbial composition of all samples.

4. Discussion

In this study, influent samples into a cosmopolitan wastewater treatment plant (WWTP) were collected over 18 consecutive days, pooled into a composite sample, and then used to compare molecular methods for detection and relative quantification of antimicrobial resistance genes (ARGs). The influent was considered because it is likely to represent the diversity in ARGs and microbiome from individual contributors to the WWTP without the complications of the biological treatment process itself, or the addition of treatment plant chemicals. In addition, it has been shown that pooling microbiome samples before DNA amplification and metagenomic sequencing is a viable approach to evaluating diversity in population-level association studies (Ray et al., 2019). We also show that the bacterial populations detected in individual influent samples could be accurately predicted by the composite pooled sample. The other main observations was that SM-Seq and qPCR yielded markedly different gene copy number (GCN) for ARGs, but normalisation of these data using z-scores demonstrated very good concordance in ARG detection and relative quantification by the two approaches. RNA sequencing identified only 32 % of specific ARGs detected by SM-Seq, and these ARGs were identified at a lower sequencing depth than that for DNA. However, when the distribution of ARGs detected by RNA sequencing were stratified by antimicrobial class against which resistance is encoded, the putative phenotypic AMR profile was not different from that determined by SM-Seq. This is because specific ARGs detected via RNA sequencing may be functional with the predicted phenotypic expression matching that of similar ARGs detected via SM-Seq in the composite DNA sample. Based on this limited dataset though, it is also apparent that a larger portion of ARGs identified via SM-Seq were possibly not transcribed into RNA transcripts at the time of sample collection.

Antimicrobial resistance gene qPCR arrays with customised amplification conditions can detect target genes with very low copy numbers (Rocha et al., 2019). This may explain the relatively higher GCN for ARGs detected by qPCR compared to SM-Seq or RNA sequencing in this study. In addition, 16 ARGs that are detectable in CARD and 18 ARGs that are detectable in MEGARes were identified by qPCR arrays in the

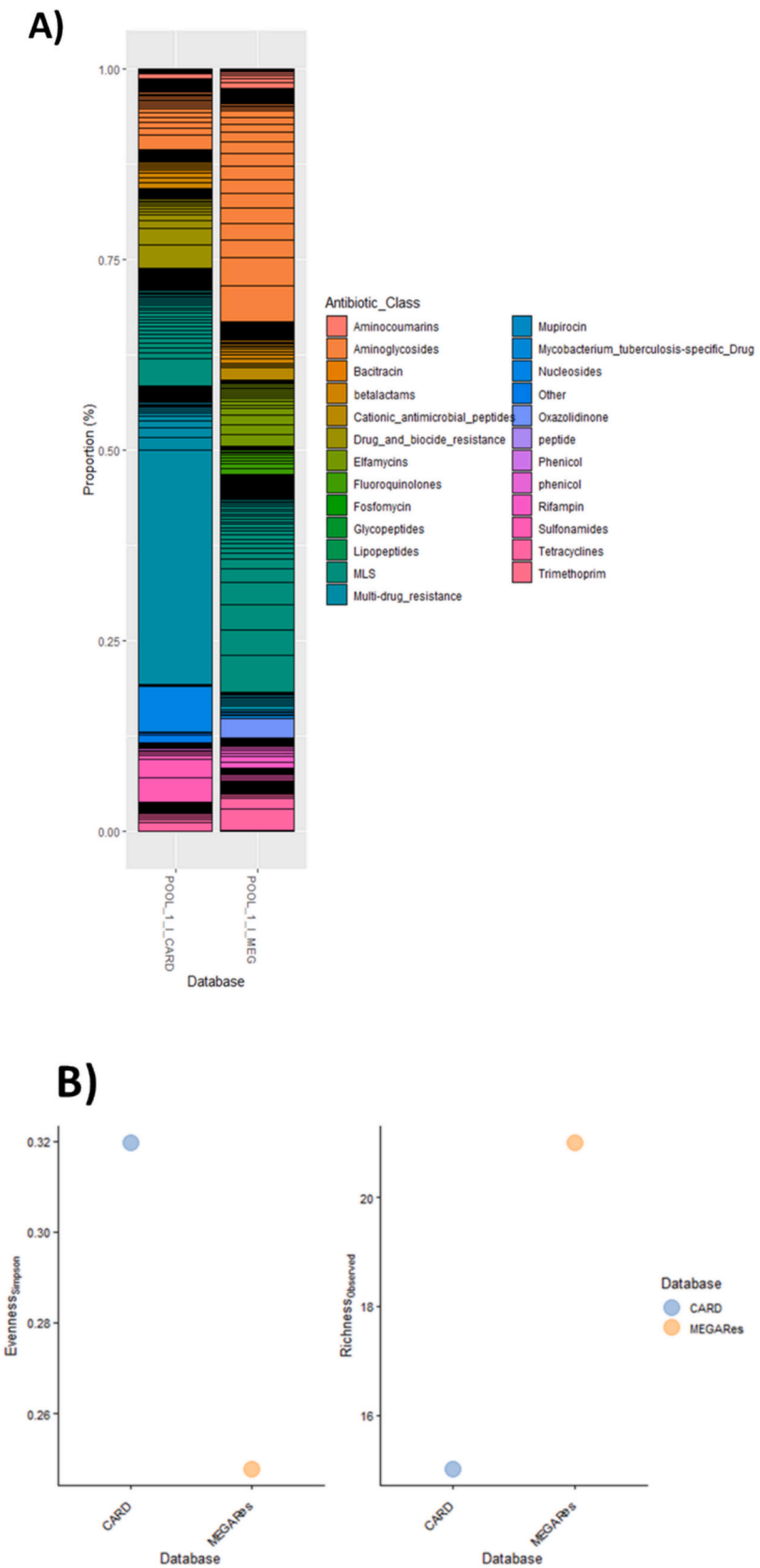

Fig. 6. Distribution of ARG sequences stratified by antimicrobial class against which they encode resistance after SM-Seq and gene annotation in the CARD and MEGARes database. (A) Relative ARG sequence abundance by antimicrobial class and (B), Alpha diversity measure of observed richness and evenness of ARG distribution as stratified by antimicrobial class.

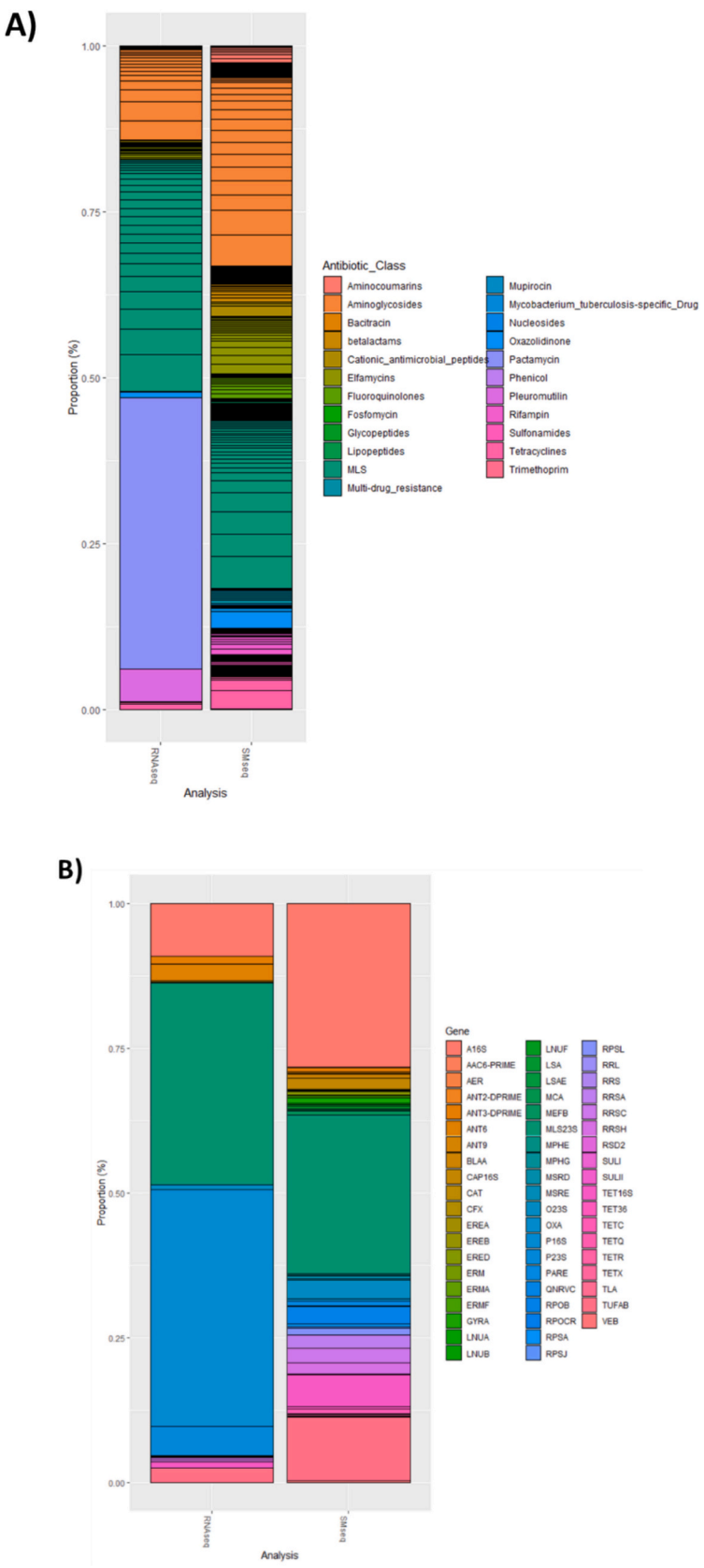

Fig. 7. Distribution of ARG sequences stratified by antimicrobial class against which they encode resistance after SM-Seq, and RNA-Seq with gene annotation in the MEGARes database. (A) Relative ARG abundance stratified by antimicrobial class in DNA and RNA sequences, (B) Relative abundance of specific ARGs in DNA and RNA sequences.

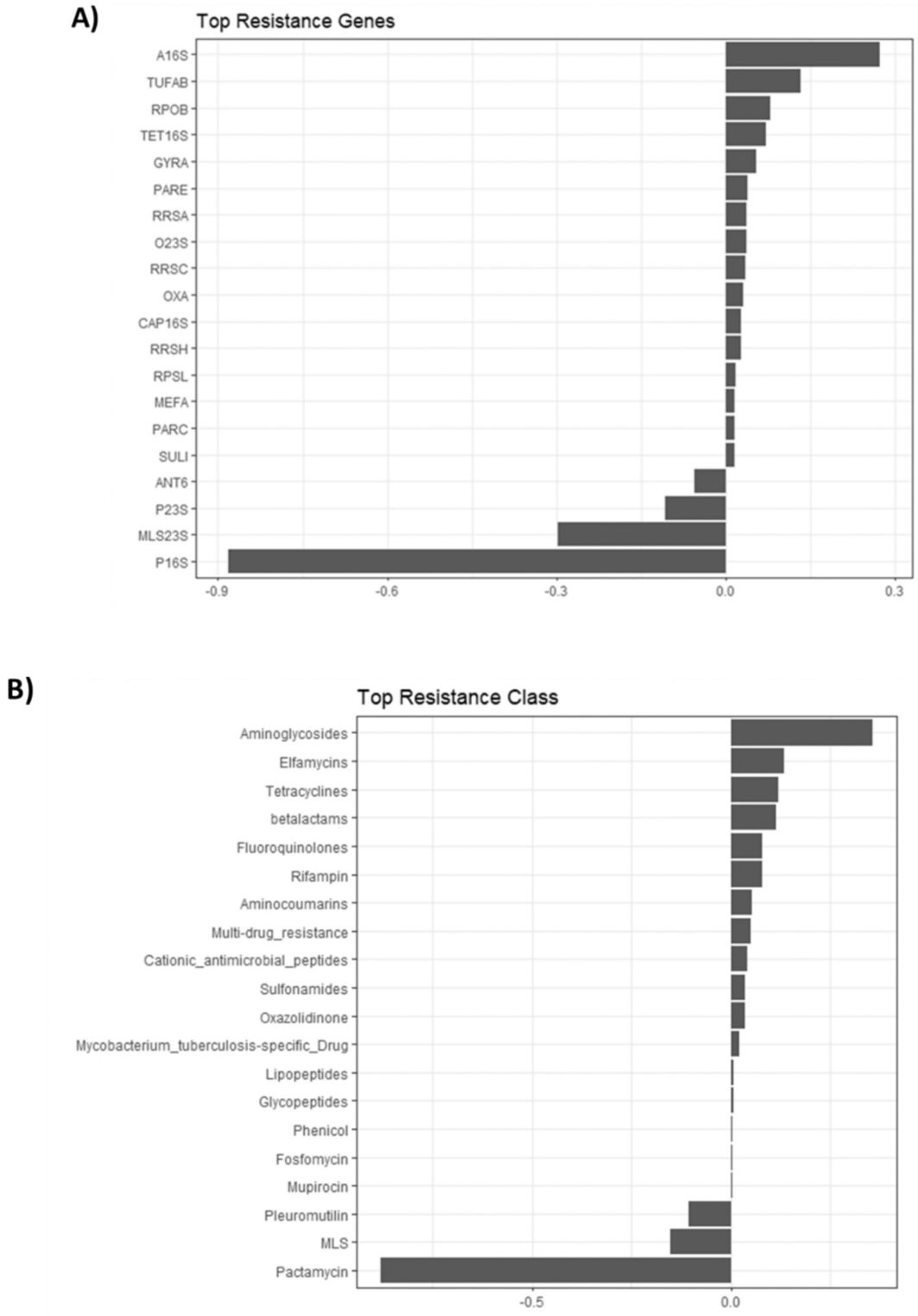

Fig. 8. Visualised model coefficients to illustrate differences in relative ARG expression in DNA compared to RNA (A) and, the antimicrobial class against which resistance is encoded (B). Values above zero indicate relatively higher abundance in DNA than RNA, and values below zero indicate relatively higher abundance in RNA than DNA.

composite sample, but were not detected by the respective SM-Seq approaches. The specific cause of this qPCR versus SM-Seq mismatch is not known. It is possible this represents false positive identification of unknown origin in qPCR since all affected ARGs were expressed at a relatively low GCN. Conversely, two ARGs including *ERMA* encoding for MLS resistance and *SME* encoding for carbapenem resistance were detected in the composite sample by SM-Seq and gene annotation in CARD or MEGARes but not via qPCR. Although not confirmed, a plausible explanation for this mismatch is that qPCR assays are highly specific with an ability to detect gene variants in a manner that is dependent on the targeted nucleotide sequences via specifically designed primers. This implies that ARG variants with subtle differences in nucleotides are

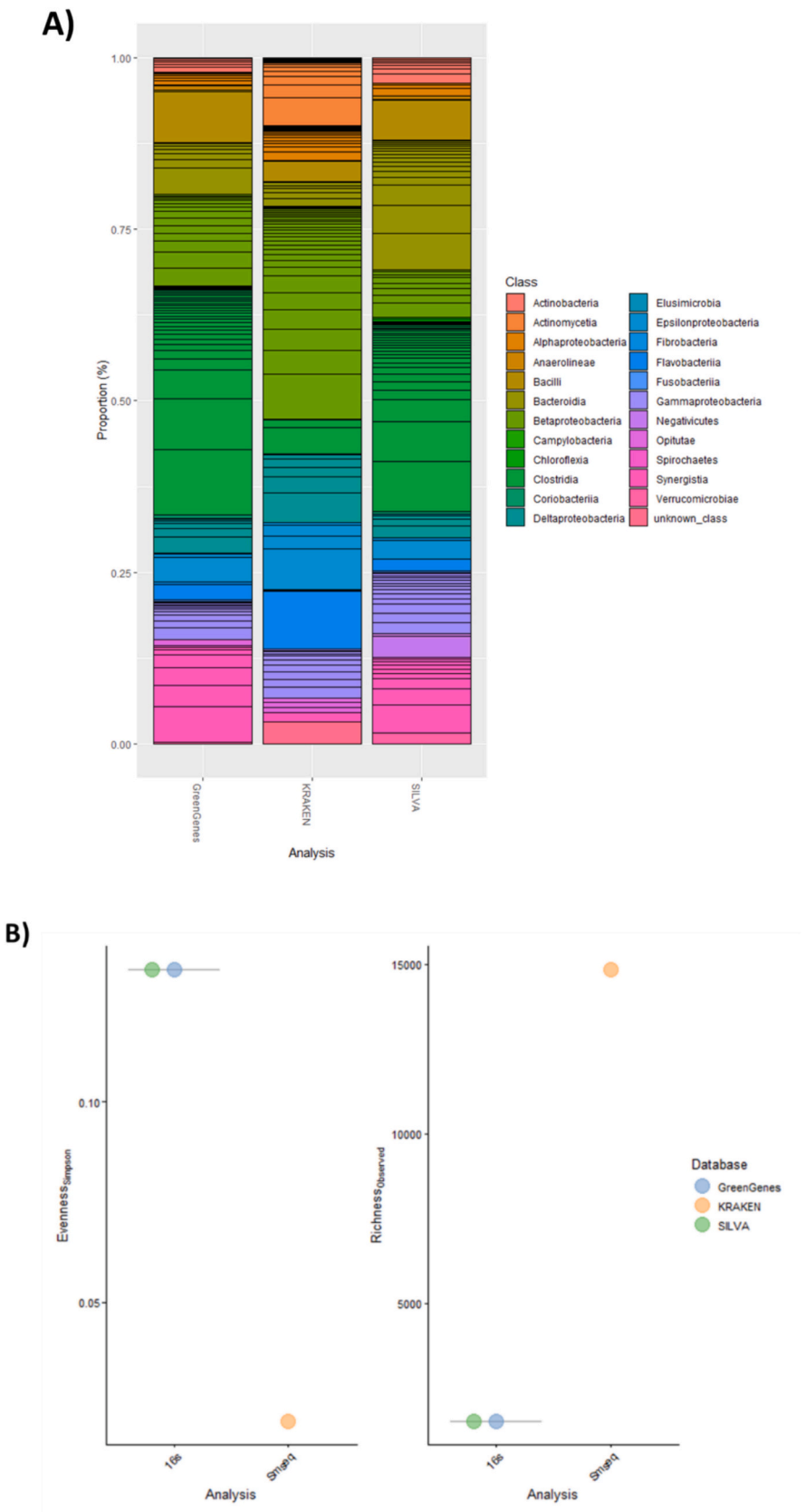

Fig. 9. A comparison of the distribution and relative proportion for classes of bacteria constituting the top 100 genera following 16S rRNA gene sequencing with taxonomic annotation in GreenGenes and SILVA, and after SM-Seq with taxonomic annotation using Kraken2 and the NCBI nucleotide database (A), and alpha diversity measure of observed richness and evenness (B). DNA was extracted from 18 influent samples, pooled into a composite sample, and then analysed.

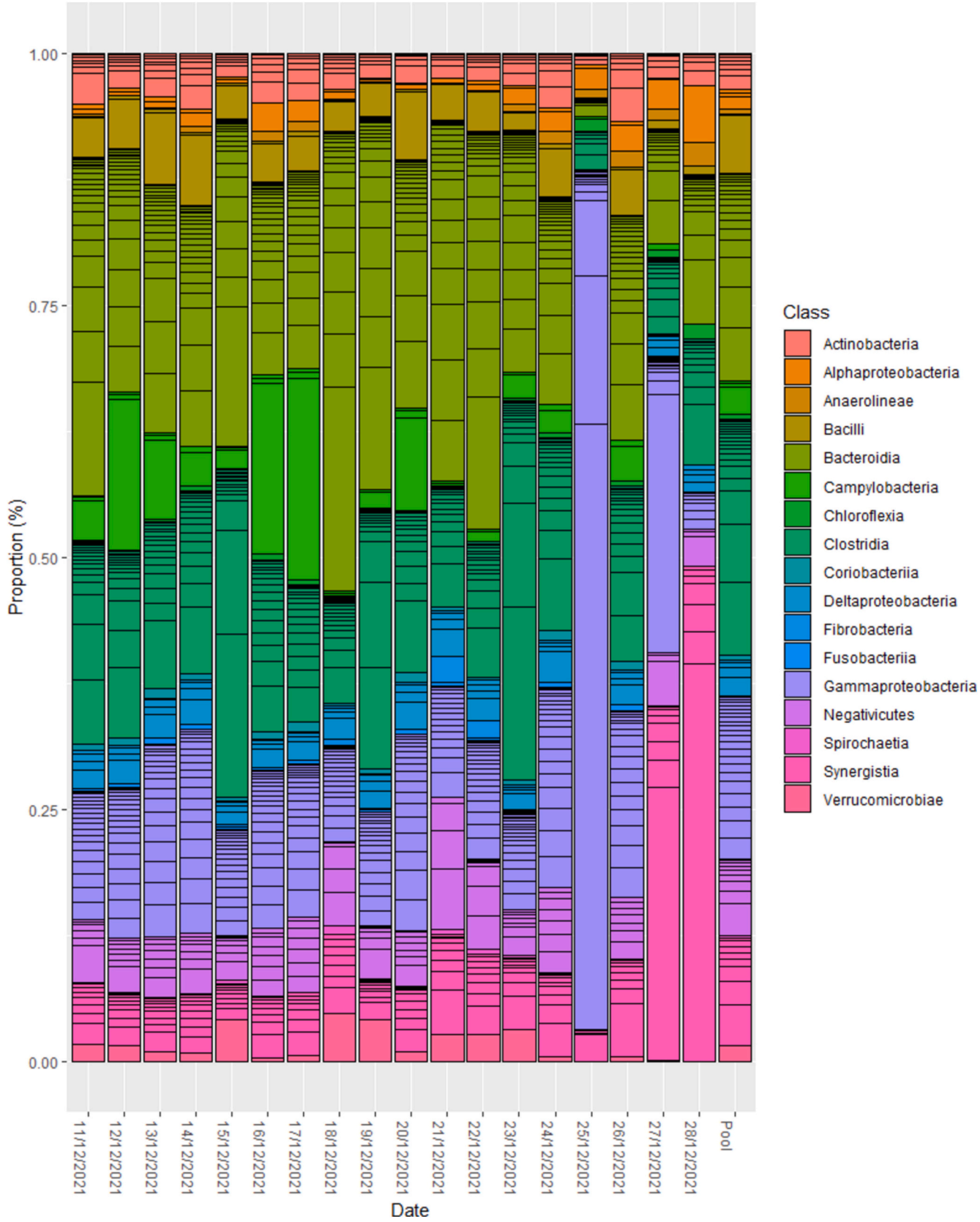

Fig. 10. Relative abundance and distribution of bacterial class in 18 individual samples and the pooled composite sample following 16S rRNA gene sequencing and bacterial class annotation using SILVA.

easily detected by reference gene sequences in metagenomic databases but could be missed by qPCR. In the MEGARes database for example, the *ERMA* gene is linked to five separate DNA sequences (Bonin et al., 2023). Furthermore, it is well known that DNA extracted from environmental samples may contain qPCR inhibitors (Stults et al., 2001). For the composite WWTP influent sample studied here however, inhibition of qPCR was unlikely because the same DNA template for ARG detection and qPCR positive controls had the same amplification efficiency. Collectively, these data show that qPCR arrays used in the current study had a limited selection of detectable ARGs, but the arrays were suitable for identification, relative quantification of ARGs in the composite influent sample. In contrast, SM-Seq provided a more comprehensive coverage of ARG detection and these methods were well suited for screening and prediction of antimicrobial resistance profiles.

Currently available databases for ARG secreening were created and are curated differently with many fundamental differences in the type and proportional representation of ARGs, gene variants or mutations (Papp and Solymosi, 2022). This study sought to compare SM-Seq and ARG annotation using either the CARD or MEGARes databases. In the composite WWTP influent sample examined here, multi-drug resistance and MLS resistance had the highest proportion of ARGs in CARD while MLS and aminoglycoside resistance had the highest proportion of ARGs in MEGARes. This is likely due to inherent differences in composition of specific ARGs in the gene pool provided in the two databases. The original CARD database was built around an ontology-driven framework which allows for only one version of an ARG sequence to exist in the database to minimize ARG allele network problem, though this could create bias for ARGs present in clinical pathogens, because of the reliance on published literature (Alcock et al., 2023). This constraint to the database may limit the identification of ARGs in environmental samples such as the CBSTP wastewater influent studied here. Furthermore, the CARD github repository (https://github.com/arpcard/rgi/blob/master/docs/rgi_bwt.rst) provides caveats that use in silico predicted allelic variants that are available in the CARD resistome. This implies that variants and prevalence data may be impacted by this network problem. Conversely, MEGARes may have multiple sequences associated with the same gene or sub-group of gene, which may be reflected in the observed higher ARG richness in the MEGARes dataset in the current study. Furthermore, AMR++ utilises bwa-mem to facilitate the read mapping of sequences which may exacerbate the network problem through redundancy in tied alignments, whereas CARD employs KMA which shows higher accuracy in bacterial genome datasets (Clausen et al., 2018). However, AMR++ utilises a cyclic form of gene annotation to minimize false identification and linking of the same gene with resistance to multiple classes of antimicrobial compounds (Papp and Solymosi, 2022; Bonin et al., 2023). Based on these differences, it is conceivable that CARD could provide a more even distribution of ARG sequences while MEGARes would provide a more comprehensive assessment of ARGs within the sample. Detailed bioinformatic analysis of the contribution of these network and pipeline differences in ARG annotation in our samples may be warranted, but this was beyond the scope of the current study.

Qualitatively, results from the current study indicate that ARG annotation in CARD or MEGARes identified different proportions of an overlapping set of ARGs that code for similar resistance patterns across the major antimicrobial classes. This is because many different bacterial genes can code for resistance against multiple antimicrobial classes creating different gene-antimicrobial class permutations that may manifest as phenotypic antibiotic resistance. A relevant example of this combination is the *ACRB* gene that codes for a multi-drug efflux pump against tetracyclines, phenicols, MLS, rifampicin and fluoroquinolones (Hobbs et al., 2012). Collectively therefore, these data suggest that CARD and MEGARes are equally suited for surveillance of known ARGs in environmental samples. For large scale studies aimed at specific ARG identification and elucidation of molecular mechanisms for AMR however, a direct comparison of CARD and MEGARes is not recommended. Furthermore, for comparative validation of detailed metagenomic analyses for ARG expression, the choice of ARG database and accompanying bioinformatic tools to be considered should be consistent.

The ARG profile identified by MEGARes and CARD databases in this composite WWTP influent sample was consistent with SM-Seq assessments of WWTP influents in some previous studies (Honda et al., 2023., Shi et al., 2023; Tiwari et al., 2024). These studies demonstrated that WWTP influents were dominated by ARGs that encode for resistance against macrolides, aminoglycosides, tetracyclines, beta-lactams, and multiple antimicrobial classes. It is worth noting however, that marked variation in the relative abundance of specific ARGs is observed across different studies and this may be attributed to qualitative differences in WWTP influent sources. The type and volume of antibiotic usage, and the presence of other factors such as the co-selection for resistance induced by metals and biocides are among the important factors that determine the microbiome and specific ARGs in WWTP influents. In the current study, a relatively high proportion of specific ARGs including *AAC6, OXA, VEB, QNRVC* were identified and these encode for resistance against aminoglycosides (*AAC*6), beta-lactam drugs (*OXA* and *VEB*), and fluoroquinolones (*QNRVC*). These antimicrobial classes are commonly used in healthcare facilities in the catchment area for the studied WWTP influent.

RNA sequencing was done to assess the putative expression and functionality of detected ARGs using the MEGARes database. MEGARes was considered for this comparison because it presented a higher ARG richness in its database compared to CARD. Fifty-nine ARGs were identified via RNA sequencing of the composite influent sample examined in this study. This however, only represented 32 % of the total number of ARGs identified by DNA sequencing and the relative abundance of genes expressed in RNA was low. This observation concurs with some previous studies. In three wastewater treatment systems in South Africa, the fate of pathogens that dominated the influent metagenomes and transcripts was found to vary markedly (Conco et al., 2022). Similarly, while resistance to most antimicrobial classes was represented in metagenomes and transcripts in activated sludge from wastewater in Taiwan, only 66 % of identified ARGs in metagenomes showed putative transcriptional activity and, some highly transcribed ARGs e.g., *APH(3)*, were not predominant in metagenomes (Liu et al., 2019b). This DNA-RNA sequencing mismatch is a common phenomenon (Pitt et al., 2020; Jia et al., 2021; Arroyo Mühr et al., 2021) and may be attributable to many factors. Primarily, RNA is composed of ribose sugars instead of deoxyribose and this makes RNA unstable compared to DNA; prone to degradation in environmental samples, by biochemical and physical elements. Degradation of messenger RNA by endonucleases and exonucleases is an important process for controlling gene expression in bacteria. It is possible that the standard protocol for nucleic acid extraction used in this study may not have sufficiently prevented RNA degradation. Secondly, RNA sequencing is only likely to represent actively expressed RNA transcripts for ARGs with a concurrent exposure to sub-lethal concentrations of corresponding antibiotics, heavy metals or nutrients at the time of sampling. For example, the level of ARG transcript abundance in wastewater effluent was found to fluctuate over time and this strongly correlated with the level of antibiotics in the effluent (Rowe et al., 2017). In *E. coli* harbouring tetracyline resistance genes, mRNA gene expression for *TETA* and *TETR* was tightly regulated by the concentration of tetracycline (Moller et al., 2016). Herein we show that the relative abundance of tetracycline resistance genes was markedly lower in RNA compared to the DNA, but determination of qualitative and quantitative features of antibiotics in samples was beyond the scope of this study. The chemical composition of the composite sample evaluated in this study may also explain why some uncommon ARGs including *P16S* encoding for pactamycin resistance, and *P23S* encoding for pleuromutilin resistance were markedly upregulated in RNA sequences. Whenever feasible therefore, real-time correlations of antimicrobial composition and concentrations with ARG transcriptomic evaluations should be considered (Drane et al., 2024; Liguori et al.,

2022). All these factors notwithstanding, ARGs with the highest abundance in both RNA and DNA sequences were *MLS23S* and *A16S*, encoding for resistance against MLS and aminoglycosides respectively. *MLS23S* facilitates a targeted alteration of 23S rRNA within the large ribosomal subunit, a target site for macrolides (Vester and Douthwaite, 2001), while *A16S* facilitates a targeted alteration of 16S rRNA within the ribosomal subunit, a target site for aminoglycosides (Lioy et al., 2014). While it is possible that the identification of *MLS23S* and *A16S* may be impacted by improper read alignment to 16S and 23S ribosomal subunits, the absence of both genes within the CARD database precluded a comparison to other alignment methods in this study. To our knowledge, these respective mutations in the 16S and 23S ribosomal subunits can occur universally and are not confined to specific organisms, and MEGARes detects these mutations. Similarly, there was no difference in the relative abundance between RNA and DNA sequences, for some ARGs encoding for phenicols, sulfonamides and multi-drug resistance. This suggests concordance in the characterisation of resistance against these specific antimicrobial classes. Similar results were observed via metagenomic and transcriptomic analyses of activated sludge from wastewater processing previously (Liu et al., 2019b).

Two methods were used to characterise the microbiome associated with the ARG profile described above; these were sequencing of 16S rRNA gene paired with bacterial class annotation in GreenGenes or SILVA databases and, SM-Seq of bacterial DNA coupled with class annotation using Kraken2 bioinformatics pipeline. Sequencing of the the 16S rRNA gene is a conventional method because it is cost effective and has large reference databases (Requeira-Iglesisa et al., 2023). This approach utilises defined operational taxonomic units that are more accurately analysed at phyla and class level but less pricisely at species level. For the composite influent sample considered here, GreenGenes and SILVA bioinformatics pipelines yielded similar microbial profiles with no differences in observed eveness or richness in the bacterial population. In contrast, SM-Seq coupled with bacterial species annotation using Kraken2 pipeline identified more bacterial classes and at much higher relative abundance with a significantly higher measure of richness in the bacterial population. This is likely due to deeper sequencing and species identification provided by SM-Seq, though considerations should be given to previous instances of taxa overexpression via this tool (Cooper et al., 2024). These data and the work elucidated by others (Durazzi et al., 2021; Lin and Liu, 2023), indicate that 16S rRNA sequencing or SM-Seq could effectively be used for bacterial population profiling at phyla and class level in surveillance studies. For more accurate speciation of bacteria in more diverse populations and identification of bacterial genes, SM-Seq would be a recommended approach. In the composite influent sample examined here, Proteobacteria, Bacteriodia and Clostridia were the predominant classes identified by both 16S rRNA sequencing and SM-Seq. This profile is very similar to that documented in influents into WWTPs in other cosmopolitan cities in the United Kingdom, Japan and China (Chau et al., 2023; Honda et al., 2023; Shi et al., 2023). This strongly suggests that influents into WWTPs in urban settings are likely to be dominated by enteric anaerobic bacteria.

There are some limitations to this study that are worth noting. Our analyses did not include an influent sample with a defined ARG and bacterial composition. As such, parameters relating to accuracy, relative sensitivity and specificity, as well as limits of detection for each of the methods was not determined. This should be considered in future studies for method standardisation. qPCR data presented here was acquired using proprietary arrays and as such, we were unable to acquire and match qPCR primers to nucleotide sequences for corresponding gene alleles detected by SM-Seq. Thus, it is not apparent whether this qPCR assay was able to detect multiple ARG alelles. While RNA sequencing may correlate with phenotypic expression of antimicrobial resistance under specific sampling conditions, bacterial culture and antimicrobial susceptibility tests that would have confirmed phenotypes were beyond the scope of this study. In addition, analysis of the composite sample was done once for MS-Seq with CARD and also once for MS-Seq with MEGARes, not replicates. This was primarily because of logistical constraints imposed by the COVID-19 pandemic when this study was conducted.

5. Conclusions

Methods that are commonly used for surveillance of the microbiome and ARGs in environmental samples are not standardised; very limited data on comparative robustness exists. We compared qPCR, SM-Seq and RNA sequensing for the qualitative and the relative quantitative detection of ARGs and bacteria in WWTP influent. SM-Seq coupled with ARG annotation in two known databases (CARD and MEGARes) provided comprehensive ARG detection that is well suited for surveillance and mapping of antimicrobial resistance profile to different antimicrobial classes. However, the detection and relative quantification of specific ARGs by this approach is limited by the size and quality of the probed ARG database. Thus, a direct comparison of CARD and MEGARes may not be recommended. It was also apparent that gene amplification using pre-selected qPCR arrays was adequate for identification, relative quantification and validation of ARGs represented in the CARD or MEGARes database. RNA sequencing identified only 32 % of the ARGs that were recognised by DNA sequencing followed by gene annotation in MEGARes. This emphasises the need to undertake phenotypic characterisations to validate AMR data determined by SM-Seq. Interestingly, all ARGs detected by RNA-seq were also detected via SM-Seq albeit at varying relative proportions. There was concordance in the detection and relative quantification of some ARGs encoding for resistance against aminoglycosides, MLS, phenicols, and sulfonamides, providing a qualitative indication of AMR patterns from RNA or DNA sequencing. Lastly, we show that sequencing of the 16S rRNA gene or SM-Seq could effectively be used for general bacterial population profiling at phyla or class level. However, SM-Seq identified more bacterial classes and at much higher relative abundance with a significantly higher measure of richness in the bacterial population. For comprehensive and precise speciation of bacteria in complex environmental samples therefore, SM-Seq would be a recommended approach.

Declaration of generative AI and AI-assisted technologies in the writing process

Generative Artificial Intelligence (AI) or AI-assisted technologies were not used in writing or generation of any of the sections of this manuscript.

Funding information

This research was funded by Townsville City Council in Queensland Australia (Grant 2256.95982.2331). Kezia Drane (KD) was a recipient of a James Cook University Postgraduate Scholarship.

Ethics statement

This work did not involve the use of animals or human participants. Influent samples into a wastewater treatment plant were collected under environmental permit G22/46568.1 issued in Queensland, Australia.

CRediT authorship contribution statement

Kezia Drane: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Roger Huerlimann:** Writing – review & editing, Validation, Supervision, Software, Resources, Methodology, Funding acquisition, Formal analysis, Data curation. **Rhondda Jones:** Writing – review & editing, Validation, Formal analysis, Data curation. **Anna Whelan:** Writing – review & editing, Resources, Project administration, Funding acquisition, Data

curation, Conceptualization. **Madoc Sheehan:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Ellen Ariel:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Robert Kinobe:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Robert Kinobe reports financial support was provided by Townsville City Council in Queensland Australia. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank Mrs. Laura Kuskopf and staff at Cleveland Bay Wastewater purification plant for assistance with with sampling, and Dr. John Otto for troubleshooting qPCR work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mimet.2024.107069.

Data availability

Data associated with this manuscript are in the James Cook University data repository (https://www.jcu.edu.au/rdim/research-data-jcu-platform) and will be made available on request.

References

Adekanmbi, A.O., Adejoba, A.T., Banjo, O.A., Saki, M., 2020. Detection of sul1 and sul2 genes in sulfonamide-resistant bacteria (SRB) from sewage, aquaculture sources, animal wastes and hospital wastewater in South-West Nigeria. Gene Rep. 20, 100742. https://doi.org/10.1016/j.genrep.2020.100742.

Alcock, B.P., Huynh, W., Chalil, R., Smith, K.W., Raphenya, A.R., Wlodarski, M.A., Edalatmand, A., Petkau, A., Syed, S.A., Tsang, K.K., 2023. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. 51 (D1), D690–D699. https://doi.org/10.1093/nar/gkac920.

Arroyo Mühr, L.S., Dillner, J., Ure, A.E., Sundstrom, K., Hultin, E., 2021. Comparison of DNA and RNA sequencing of total nucleic acids from human cervix for metagenomics. Sci. Rep. 11 (1), 18852. https://doi.org/10.1038/s41598-021-98452-4.

Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., Huttley, G. A., Caporaso, J.G., 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6 (1), 90. https://doi.org/10.1186/s40168-018-0470-z.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J.E., Bittinger, K., Brejnrod, A., Brislawn, C.J., Brown, C.T., Callahan, B.J., Caraballo-rodrigues, A.M., Chase, J., Cope, E.K., Da Silva, R., Dierner, C., Dorrestein, P.C., Douglas, G.M., Caporaso, J.G., 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37 (8), 852–857. https://doi.org/10.1038/s41587-019-0209-9.

Bonetta, S., Di Cesare, A., Pignata, C., Sabationo, R., Macri, M., Corno, G., Panizzolo, M., Bonetta, S., Carraro, E., 2023. Occurrence of antibiotic-resistant bacteria and resistance genes in the urban water cycle. Environ. Sci. Pollut. Res. Int. 30 (12), 35294–35306. https://doi.org/10.1007/s11356-022-24650-w.

Bonin, N., Doster, E., Worley, H., Pinnell, L.J., Bravo, J.E., Ferm, P., Marini, S., Prosperi, M., Noyes, N., Morley, P.S., 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. Nucleic Acids Res. 51 (D1), D744–D752. https://doi.org/10.1093/nar/gkac1047.

Boolchandani, M., D'souza, A.W., Dantas, G., 2019. Sequencing-based methods and resources to study antimicrobial resistance. Nat. Rev. Genet. 20 (6), 356–370. https://doi.org/10.1038/s41576-019-0108-4.

Callahan, B.J., Mcmurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 13 (7), 581–583. https://doi.org/10.1038/nmeth.3869.

Chau, K.K., Goodall, T., Bowes, M., Easterbrook, K., Brett, H., Hughes, J., Crook, D.W., Read, D.S., Walker, A.S., Stoesser, N., 2023. High-resolution characterization of short-term temporal variability in the taxonomic and resistome composition of wastewater influent. Microb. Genom. 9 (5), mge000983. https://doi.org/10.1099/mgen.0.000983.

Christgen, B., Yang, Y., Ahammad, S., Li, B., Rodriquez, D.C., Zhang, T., Graham, D.W., 2015. Metagenomics shows that low-energy anaerobic− aerobic treatment reactors reduce antibiotic resistance gene levels from domestic wastewater. Environ. Sci. Technol. 49 (4), 2577–2584. https://doi.org/10.1021/es505521w.

Clausen, P.T.L., Aarestrup, F.M., Lund, O., 2018. Rapid And Precise Alignment Of Raw Reads Against Redundant Databases With KMA. BMC Bioinformatics 19 (1), 307. https://doi.org/10.1186/s12859-018-2336-6.

Conco, T., Kumari, S., Awolusi, O.O., Allam, M., Ismail, A., Stenstrom, T.A., Bux, F., 2022. Profiling of emerging pathogens, antibiotic resistance genes and mobile genetic elements in different biological wastewater treatment plants. J. Environ. Chem. Eng. 10 (3), 107596. https://doi.org/10.1016/j.jece.2022.107596.

Cooper, A.L., Low, A., Wong, A., Tamber, S., Blais, B.W., Carrillo, C.D., 2024. Modeling the limits of detection for antimicrobial resistance genes in agri-food samples: a comparative analysis of bioinformatic tools. BMC Microbiol. 24 (1), 31. https://doi.org/10.1186/s12866-023-03148-6.

Desantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72 (7), 5069–5072.

Drane, K., Sheehan, M., Whelan, A., Ariel, E., Kinobe, R., 2024. The role of wastewater treatment plants in dissemination of antibiotic resistance: source, measurement, removal and risk assessment. Antibiotics 13 (7), 668. https://doi.org/10.3390/antibiotics13070668.

Durazzi, F., Sala, C., Castellani, G., Manfreda, G., Remondini, D., De Cesare, A., 2021. Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. Sci. Rep. 11 (1), 3030. https://doi.org/10.1038/s41598-021-82726-y.

Gaviria-Figueroa, A., Preisner, E.C., Hoque, S., Feigley, C.E., Norman, R.S., 2019. Emission and dispersal of antibiotic resistance genes through bioaerosols generated during the treatment of municipal sewage. Sci. Total Environ. 686, 402–412. https://doi.org/10.1016/j.scitotenv.2019.05.454.

Gholipour, S., Shamsizadeh, Z., Halabowski, D., Gwenzi, W., Nikaeen, M., 2024. Combating antibiotic resistance using wastewater surveillance: Significance, applications, challenges, and future directions. Sci. Total Environ. 908, 168056. https://doi.org/10.1016/j.scitotenv.2023.168056.

Guo, J., Li, J., Chen, H., Bond, P.L., Yuan, Z., 2017. Metagenomic analysis reveals wastewater treatment plants as hotspots of antibiotic resistance genes and mobile genetic elements. Water Res. 123, 468–478. https://doi.org/10.1016/j.watres.2017.07.002.

Hendriksen, R.S., Munk, P., Njage, P., Van Bunnik, B., Mcnally, L., Lukjancenko, O., Röder, T., Nieuwenhuijse, D., Pedersen, S.K., Kjeldgaard, J., Kaas, R.S., Clausen, P.T. L.C., Vogt, J.K., Leekitcharoenphon, P., Van De Schans, M.G.M., Zuidema, T., De Rod 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., 2019. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. Nat. Commun. 10 (1), 1124. https://doi.org/10.1038/s41467-019-08853-3.

Hobbs, E.C., Yin, X., Paul, B.J., Astarita, J.L., Storz, G., 2012. Conserved small proteins associates with the multdrug efflux pump AcrB and differentially affects antibiotic resistance. Proc. Natl. Acad. Sci. USA 109 (41), 16696–16701. https://doi.org/10.1073/pnas.1210093109.

Honda, R., Matsuura, N., Sorn, S., Asakura, S., Morinaga, Y., Van Huy, T., Sabar, M.A., Masakke, Y., Hara-Yamamura, H., Watanabe, T., 2023. Transition of antimicrobial resistome in wastewater treatment plants: impact of process configuration, geographical location and season. npj Clean Water 6, 46. https://doi.org/10.1038/s41545-023-00261-x.

Jia, S., Li, T., Zhang, X.-X., 2021. Integrated metagenomic and metatranscriptomic of ultraviolet disinfection effects on antibiotic resistance genes and bacterial communities during wastewater treatment. Ecotoxicology 30 (8), 1610–1619. https://doi.org/10.1007/s10646-020-02313-1.

Kruerger, F., 2015. Trim Galore. A Wrapper around Cutadapt and FastQC to consistently apply adapter abd quality trimming to FastQ files, with extra functionality for RRBS data. https://github.com/FelixKrueger/TrimGalore (Last accessed September 30, 2024).

Larsson, D.G.J., Flach, C.-F., 2022. Antibiotic resistance in the environment. Nat. Rev. Microbiol. 20 (5), 257–269. https://doi.org/10.1038/s41579-021-00649-x.

Liguori, K., Keenum, I., Davis, B.C., Calarco, J., Milligan, E., Harwood, V.J., Pruden, A., 2022. Antimicrobial resistance monitoring of water environments: A framework for standardised methods and quality control. Environ. Sci. Technol. 56, 9149–9160. https://doi.org/10.1021/acs.est.1c08918.

Lin, T., Liu, W., 2023. Validation of 16S rRNA gene sequencing and metagenomics for evaluating microbial immigration in a methanogenic bioreactor. Water Res. 243, 120358. https://doi.org/10.1016/j.watres.2023.120358.

Lioy, V.S., Goussard, S., Guerineau, V., Yoon, E.J., Courvalin, P., Galimand, M., Grillot-Courvalin, C., 2014. Aminoglycoside resistance 16S rRNA methyltransferases block endogenous methylation, affect translation efficiency and fitness of the host. RNA 20 (3), 382–391. https://doi.org/10.1261/rna.042572.113.

Liu, L., Su, J.-Q., Guo, Y., Wilkinson, D.M., Liu, Z., Zhu, Y.-G., Yang, J., 2018. Large-scale biogeographical patterns of bacterial antibiotic resistome in the waterbodies of China. Environ. Int. 117, 292–299. https://doi.org/10.1016/j.envint.2018.05.023.

Liu, X., Xiao, P., Guo, Y., Liu, L., Yang, J., 2019a. The impacts of different high-throughput profiling approaches on the understanding of bacterial antibiotic resistance genes in a freshwater reservoir. Sci. Total Environ. 693, 133585. https://doi.org/10.1016/j.envint.2018.05.023.

Liu, Z., Klumper, U., Liu, Y., Yang, Y., Wei, Q., Lin, J., Gu, J., Li, M., 2019b. Metagenomic and metatranscriptomic analyses reveal activity and hosts of antibiotic resistance genes in activated sludge. Environ. Int. 129, 208–220. https://doi.org/10.1016/j.envint.2019.05.036.

Looft, T., Johnson, T.A., Allen, H.K., Bayles, D.O., Alt, D.P., Stedtfeld, R.D., Sul, W.J., Stedtfeld, T.M., Chai, B., Cole, J.R., Hashsham, S.A., Tiedje, J.M., Stanton, T.B., 2012. In-feed antibiotic effects on the swine intestinal microbiome. Proc. Natl. Acad. Sci. USA 109 (5), 1691–1696. https://doi.org/10.1073/pnas.1120238109.

Lu, J., Breitwieser, F.P., Thielen, P., Salzberg, S.L., 2017. Bracken: estimating species abundance in metagenomics data. PeerJ Comput. Sci. 3, e104. https://doi.org/10.7717/peerj-cs.104.

Majeed, H.J., Riquelme, M.V., Davis, B.C., Gupta, S., Angeles, L., Aga, D.S., Garner, E., Pruden, A., Vikesland, P.J., 2021. Evaluation of metagenomic-enabled antibiotic resistance surveillance at a conventional wastewater treatment plant. Front. Microbiol. 12, 657954. https://doi.org/10.3389/fmicb.2021.657954.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. 17 (1), 10–12. https://doi.org/10.14806/ej.17.1.200.

Mason, A., Foster, D., Bradley, P., Golubchik, T., Doumith, M., Gordon, N.C., Pichon, B., Iqbal, Z., Staves, P., Crook, D., Walker, A.S., Kearns, A., Peto, T., 2018. Accuracy of different bioinformatics methods in detecting antibioc resistance and virulence factors from *Staphylococcus aureus* whole-genome sequences. J. Clin. Microbiol. 56 (9). https://doi.org/10.1128/JCM.01815-17 e01818-17.

Mcmurdie, P.J., Holmes, S., 2012. Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data. Pac. Symp. Biocomput. 17, 235–246. https://doi.org/10.1142/9789814366496_0023.

Moller, T.S.B., Overgaard, M., Nielsen, S.S., Bortolaia, V., Sommer, M.O.A., Guardabassi, L., Olsen, J.E., 2016. Relation between *tetR* and *tetA* expression in tetracycline resistant *Escherichia coli*. BMC Microbiol. 16, 39. https://doi.org/10.1186/s12866-016-0649-z.

Muurinen, J., Karkman, A., Virta, M., 2017. High throughput method for analyzing antibiotic resistance genes in wastewater treatment plants. In: Antimicrobial Resistance in Wastewater Treatment Processes, 253, p. 262. https://doi.org/10.1002/9781119192428.ch14.

Odom, A.R., Faits, T., Castro-Nallar, E., Crandall, K.A., Johnson, W.E., 2023. Metagenomic profiling pipelines improve taxonomic classification for 16S amplicon sequencing data. Sci. Rep. 13 (1), 13957. https://doi.org/10.1038/s41598-023-40799-x.

Papp, M., Solymosi, N., 2022. Review and comparison of antimicrobial resistance gene databases. Antibiotics 11 (3), 339. https://doi.org/10.3390/antibiotics11030339. PMID: 35326803.

Pitt, M.E., Nguyen, S.H., Duarte, T.P.S., Teng, H., Blaskovich, M.A.T., Cooper, M.A., Coin, L.J.M., 2020. Evaluating the genome and resistome of extensively drug-resistant *Klebsiella pneumoniae* using native DNA and RNA Nanopore sequencing. Gigascience 9 (2), giaa002. https://doi.org/10.1093/gigascience/giaa002.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596. https://doi.org/10.1093/nar/gks1219.

Ray, K.J., Cotter, S.Y., Arzika, A.M., Kim, J., Boubacar, N., Zhou, Z., Zhong, L., Porco, T. C., Keenan, J.D., Lietman, T.M., Doan, T., 2019. High-throughput sequencing of pooled samples to determine community-level microbiome diversity. Ann. Epidemiol. 39, 63–68. https://doi.org/10.1016/j.annepidem.2019.09.002.

Requeira-Iglesisa, A., Balsa-Castro, C., Blanco-Pintos, T., Tomas, I., 2023. Critical review of 16S rRNA gene sequencing workflow in microbiome studies: From primer selection to advanced data analysis. Mol Oral Microbiol 38950, 347–399. https://doi.org/10.1111/omi.12434.

Rocha, J., Fernandes, T., Riquelme, M.V., Zhu, N., Pruden, A., Maanaia, C.M., 2019. Comparison of culture- and quantitative PCR-based indicators of antibiotic resistance in wastewater, recycled water, and tap water. Int. J. Environ. Res. Public Health 16 (21), 4217. https://doi.org/10.3390/ijerph16214217.

Rowe, W.P., Baker-Austin, C., Verner-Jeffreys, D.W., Ryan, J.J., Micallef, C., Maskell, D. J., Pearce, G.P., 2017. Overexpression of antibiotic resistance genes in hospital effluents overtime. J. Antimicrob. Chemother. 72 (6), 1617–1623. https://doi.org/10.1093/jac/dkx017.

Shi, B., Zhao, R., Su, G., Liu, B., Liu, W., Xu, J., Li, Q., Meng, J., 2023. Metagenomic surveillance of antibiotic resistome in influent and effluent of wastewater treatment plants located on the Qinghai-Tibetan Plateau. Sci. Total Environ. 870, 162031. https://doi.org/10.1016/j.scitotenv.2023.162031.

Stults, J.R., Snoeyenbos-West, O., Methe, B., Lovely, D.R., Chandler, D.R., 2001. Application of the 5′ fluorogenie exonuclease assay (TaqMan) for quantitative ribosomal DNA and rRNA analysis in sediments. Appl. Environ. Microbiol. 67, 2781–2789. https://doi.org/10.1128/AEM.67.6.2781-2789.2001.

Tiwari, A., Kurittu, P., Al-Mustapha, A.I., Heljanko, V., Johansson, V., Thakali, O., Mishra, S.K., Lehto, K.M., Lipponen, A., Oikarinen, S., Pitkanen, T., Wastpan Study Group, Heikinheimo, A., 2022. Wastewater surveillance of antibiotic-resistant bacterial pathogens: A systematic review. Front. Microbiol. 13, 977106. https://doi.org/10.3389/fmicb.2022.977106.

Tiwari, A., Krolicka, A., Tran, T.T., Raisanen, K., Asmundsdottir, A.M., Wikmark, O.-G., Lood, R., Pitkanen, T., 2024. Antibiotic resistance monitoring in wastewater in the Nordic countries: A systematic review. Environ. Res. 246, 118052. https://doi.org/10.1016/j.envres.2023.118052.

Vester, B., Douthwaite, S., 2001. Macrolide resistance conferred by base substitutions in 23S rRNA. Antimicrob. Agents Chemother. 45 (1), 1–12. https://doi.org/10.1128/AAC.45.1.1-12.2001.

Wang, N., Yang, X., Jiao, S., Zhang, J., Ye, B., Gao, S., 2014. Sulfonamide-resistant bacteria and their resistance genes in soils fertilized with manures from Jiangsu Province, Southeastern China. PLoS One 9 (11), e112626. https://doi.org/10.1371/journal.pone.0112626.

Wickham, H., 2011. ggplot2. WIREs Comp. Stat. 3, 180–185. https://doi.org/10.1002/wics.147.

Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. Genome Biol. 20 (1), 257. https://doi.org/10.1186/s13059-019-1891-0.