

UC Irvine

UC Irvine Previously Published Works

Title

Longitudinal metatranscriptomic sequencing of Southern California wastewater representing 16 million people from August 2020-21 reveals widespread transcription of antibiotic resistance genes

Permalink

<https://escholarship.org/uc/item/31w6s89q>

Authors

Rothman, Jason A
Saghir, Andrew
Chung, Seung-Ah
et al.

Publication Date

2023-02-01

DOI

10.1016/j.watres.2022.119421

Data Availability

The data associated with this publication are in the supplemental files.

Peer reviewed



Longitudinal metatranscriptomic sequencing of Southern California wastewater representing 16 million people from August 2020–21 reveals widespread transcription of antibiotic resistance genes

Jason A. Rothman^{a,*}, Andrew Saghir^a, Seung-Ah Chung^b, Nicholas Boyajian^a, Thao Dinh^a, Jinwoo Kim^a, Jordan Oval^a, Vivek Sharavanan^a, Courtney York^a, Amity G. Zimmer-Faust^c, Kylie Langlois^c, Joshua A. Steele^c, John F. Griffith^c, Katrine L. Whiteson^{a,*}

^a Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine, CA, United States of America

^b Genomics High-Throughput Facility, Department of Biological Chemistry, University of California, Irvine, Irvine, CA, United States of America

^c Southern California Coastal Water Research Project, Costa Mesa, CA, United States of America

ARTICLE INFO

Keywords:

Wastewater
Antimicrobial resistance
Metatranscriptomics
Microbial ecology
Environmental microbiology

ABSTRACT

Municipal wastewater provides a representative sample of human fecal waste across a catchment area and contains a wide diversity of microbes. Sequencing wastewater samples provides information about human-associated and medically important microbial populations, and may be useful to assay disease prevalence and antimicrobial resistance (AMR).

Here, we present a study in which we used untargeted metatranscriptomic sequencing on RNA extracted from 275 sewage influent samples obtained from eight wastewater treatment plants (WTPs) representing approximately 16 million people in Southern California between August 2020 – August 2021. We characterized bacterial and viral transcripts, assessed metabolic pathway activity, and identified over 2,000 AMR genes/variants across all samples. Because we did not deplete ribosomal RNA, we have a unique window into AMR carried as ribosomal mutants. We show that AMR diversity varied between WTPs (as measured through PERMANOVA, $P < 0.001$) and that the relative abundance of many individual AMR genes/variants increased over time (as measured with MaAsLin2, $P_{adj} < 0.05$). Similarly, we detected transcripts mapping to human pathogenic bacteria and viruses suggesting RNA sequencing is a powerful tool for wastewater-based epidemiology and that there are geographical signatures to microbial transcription. We captured the transcription of gene pathways common to bacterial cell processes, including central carbon metabolism, nucleotide synthesis/salvage, and amino acid biosynthesis. We also posit that due to the ubiquity of many viruses and bacteria in wastewater, new biological targets for microbial water quality assessment can be developed.

To the best of our knowledge, our study provides the most complete longitudinal metatranscriptomic analysis of a large population's wastewater to date and demonstrates our ability to monitor the presence and activity of microbes in complex samples. By sequencing RNA, we can track the relative abundance of expressed AMR genes/variants and metabolic pathways, increasing our understanding of AMR activity across large human populations and sewer sheds.

1. Introduction

Wastewater harbors a wide diversity of microorganisms and represents the collective waste of human activity across a sewershed (Newton and McClary, 2019). Over 300 km³ of wastewater is produced globally, of which most is channeled into wastewater treatment plants (WTPs) for

biological and chemical processing (Lu et al., 2018). As a heterogeneous mixture, wastewater has been shown to contain microbial communities that vary depending on sampling location, time of year, industry, agriculture, and the health of the served human population (Cantalupo et al., 2011; Edwards et al., 2019; McLellan et al., 2010; Symonds et al., 2009; Wu et al., 2019). As a result, the microbial water quality of

* Co-Corresponding authors at: University of California, Irvine, Irvine, CA, 92697.

E-mail addresses: rothmanj@uci.edu (J.A. Rothman), katrine@uci.edu (K.L. Whiteson).

<https://doi.org/10.1016/j.watres.2022.119421>

Received 10 August 2022; Received in revised form 22 November 2022; Accepted 23 November 2022

Available online 24 November 2022

0043-1354/© 2022 Elsevier Ltd. All rights reserved.

wastewater can be a useful indicator of an area's biological contamination, with outbreaks of several diseases corresponding to increased wastewater titers of pathogenic etiological agents (Hellmér et al., 2014; Manor et al., 1999; Rothman et al., 2021; Wu et al., 2020). The microbial ecology of wastewater is an important topic, with many studies characterizing the microbes present through culturing, PCR- and sequencing-based methods. These studies generally rely on targeting specific pathogens or metagenomic shotgun DNA sequencing (Hubeny et al., 2022; Jankowski et al., 2022; Kitajima et al., 2018; Martínez-Puchol et al., 2020). While useful, these studies are unable to capture microbial transcription, which provides information about active microbial processes, instead of the genomic potential of wastewater. Moreover, as many important human and crop/livestock pathogens are RNA viruses (Amoah et al., 2020; Bibby and Peccia, 2013; Symonds et al., 2009), we can monitor the presence and spread of *Ribovira* through untargeted metatranscriptomics. Wastewater RNA sequencing can uncover active microbial interactions and metabolic networks, which may inform us of the public and environmental health of the areas served by a given sewage system (Brumfield et al., 2022; Crits-Christoph et al., 2021; Li et al., 2022; Rothman et al., 2021).

Wastewater-based epidemiology (WBE) can inform public health about the presence of pathogens in a population without needing to test individuals in healthcare settings (Bivins et al., 2020; Sims and Kasprzyk-Hordern, 2020). Health agencies have used WBE to detect the presence of human pathogens such as norovirus, polio, SARS coronaviruses, and a variety of bacteria and protists (Hellmér et al., 2014; Manor et al., 1999; Rothman et al., 2021; Wu et al., 2020). For example, WBE has been heavily used to track and monitor the abundance and spread of SARS-CoV-2 during the ongoing COVID-19 pandemic at various population levels (Achak et al., 2021; Karthikeyan et al., 2021; Nemudryi et al., 2020; Peccia et al., 2020; Rothman et al., 2021; Wu et al., 2020). Furthermore, as disease case counts change longitudinally, multiple time points and RNA sequencing are useful to track not only the presence, but the activity of microorganisms which may provide additional information about pathogens over longer time periods (Faust et al., 2015; Joseph et al., 2019; Marcelino et al., 2019; Nemudryi et al., 2020). Lastly, by broadly sequencing RNA, we may be able to discover new targets for microbial water quality assays in order to detect and monitor for sewage contamination of the environment and water sources (Cao et al., 2015; Farkas et al., 2019; Jiang et al., 2022; Kitajima et al., 2018; Zimmer-Faust et al., 2021).

Antimicrobial resistance (AMR) is a worldwide concern that inhibits effective treatment of disease and increases healthcare burden and morbidity of infections (World Health Organization, 2021). Wastewater contains a complex diversity of AMR genes, which allows for horizontal gene transfer (HGT) of mobile genetic elements between antimicrobial resistant organisms and those species or strains that are currently susceptible to antimicrobial therapies (Joseph et al., 2019; Ju et al., 2019; Sims and Kasprzyk-Hordern, 2020). As AMR and HGT are important to monitor for public and agricultural health, many studies have employed sequencing and targeted PCR-based technologies to assay the AMR genomic content of wastewater (Karkman et al., 2018). While useful, these studies typically rely on DNA-based technologies which cannot measure the transcriptional activity of these genes or the organisms that harbor AMR, and may better indicate the severity and abundance of antimicrobial resistant infections across a population (de Nies et al., 2021; Ju et al., 2019; Marcelino et al., 2019). By employing RNA-sequencing, we are better able to understand the disease ecology and AMR activity of wastewater-inhabiting organisms and those deposited through the waste stream, and the specific mutations that cause AMR (Alcock et al., 2020). Lastly, through careful sampling, changes in AMR transcription can be tracked over time, likely providing finer-scale information about the severity and seasonality of AMR infections (Langford et al., 2020; Rose et al., 2021).

Studying wastewater microbial ecology and tracking the activity of disease-associated microbes and AMR is vital to public health and

environmental monitoring. In this study, we used metatranscriptomic sequencing to characterize the RNA world of 275 samples across eight wastewater treatment plants (WTPs) representing approximately 16 million people across Southern California. We investigated several lines of inquiry: First, what is the transcriptomic diversity of microorganisms in Southern California wastewater, and does it vary longitudinally? Second, what AMR genes are being actively transcribed in wastewater? Third, are there conserved biochemical pathways across wastewater, and does this metabolic potential vary? Lastly, are there largescale patterns of microbial transcription in Southern California's wastewater, and is there a temporal component to any of these patterns?

2. Materials and methods

2.1. Sample collection

We previously reported the sample collection and handling procedure in Rothman et al., 2021 (Rothman et al., 2021), and note that the viromes of 94 samples were previously reported in that study. Briefly, we collected 275 1-liter 24-hour composite influent wastewater samples by autosampler at eight WTPs across Southern California between August 2020 – August 2021 (Table 1, Fig. 1). We aliquoted and stored 50 mL of sample at 4 °C until RNA extraction.

2.2. RNA extraction and sequencing library preparation

We used a protocol based on Crits-Christoph 2021 (Crits-Christoph et al., 2021) and Wu et al. 2020 (Wu et al., 2020), in which we pasteurized 50 mL of influent wastewater in a 65 °C water bath for 90 min, then filtered samples through a sterile 0.22-µm filter (VWR, Radnor, PA). We then centrifuged the sample at 3000 xg through 10-kDa

Table 1

Descriptions of the experiment sampling scheme and relevant information about each WTP.

Wastewater Treatment Plant	Number of Samples	Date Span	Approximate Inflow (Million Gallons/Day)	Approximate Population Served
Escondido Hale Avenue Resource Recovery Facility	45	August 3 2020, – July 19 2021,	14	190,000
Hyperion Water Reclamation Plant	92	August 11 2020, – July 29 2021,	275	4000,000
Joint Water Pollution Control Plant	15	August 11 2020, – November 17 2020,	400	4800,000
North City Water Reclamation Plant	7	August 14 2020, – November 6 2020,	30	1400,000
Orange County Reclamation Plant #1	17	August 12 2020, – December 21 2020,	140	2600,000
Point Loma Water Treatment Plant	77	August 13 2020, – August 3 2021,	175	2200,000
San Jose Creek Water Reclamation Plant	15	August 12 2020, – November 18 2020,	100	1000,000
South Bay Water Reclamation Plant	7	August 13 2020, – November 5 2020,	15	107,000

Amicon filters (MilliporeSigma, Burlington, MA) to concentrate the samples' volume to < 500 μ L and stored the concentrate at -80°C until RNA extraction. We then used an Invitrogen PureLink RNA Mini Kit with added DNase step (Invitrogen, Waltham, MA) following the manufacturer's instructions to extract RNA and stored the resulting RNA at -80°C until library preparation.

The University of California Irvine Genomics High Throughput Facility (GHTF) handled all library preparation steps. Briefly, the GHTF used the Illumina RNA Prep for Enrichment kit (Illumina, San Diego, CA) on each RNA sample, then sequenced the paired end libraries as 2×100 bp or 2×150 bp (supplemental file SF1) with an S4 300 cycle kit on an Illumina NovaSeq 6000 over four batches.

2.3. Bioinformatics and data processing

We received the data from the GHTF as demultiplexed FASTQ files and used the UCI High Performance Community Computing Cluster for data processing. We used BBTools v38.87 "bbduk" (Bushnell, 2014) to remove Illumina adapters, low-quality bases (Phred score < 10), and primers, then removed PCR duplicates with BBTools "dedupe." After deduplication, we removed reads mapping to the human genome (hg38) with Bowtie2 v2.4.1 with the default settings (Langmead and Salzberg, 2012), then used Kraken2 v2.0.7 (Wood et al., 2019) and Bracken v2.6.1 (Lu et al., 2017) with databases built with the NCBI RefSeq database of bacteria, archaea, and viruses (January 2021), to taxonomically classify reads with default settings and Phred score > 20. We then tabulated these reads and used these tables for downstream diversity analysis and reported raw read numbers ((supplemental file SF1 and Dryad dataset (<https://doi.org/10.7280/D11Q30>) (Rothman et al., 2022)). Lastly, we used EukDetect v1.3 (Lind and Pollard, 2021) to attempt classification of eukaryotes in our samples.

For community analyses, we normalized the transcript reads into within-sample relative abundances in R, removed reads corresponding to less than 0.01% relative abundance, then calculated Shannon Diversity indices and Bray-Curtis dissimilarity matrices with the R package "vegan" (Oksanen et al., 2017). We generated nonmetric multidimensional scaling (NMDS) ordinations, then tested the diversity metrics for significant differences with Kruskal-Wallis tests (alpha diversity) and Adonis PERMANOVA (beta diversity) with "vegan." We assessed the relationship of diversity with time with linear mixed effects models (lmer) in the R package "lmerTest" using WTP and sequencing batch as random effects (Kuznetsova et al., 2017), and plotted all diversity analyses with "ggplot2" (Wickham, 2009), "ggrepel" (Slowikowski, 2018), and "patchwork" (Pedersen, 2020). Because we collected samples from Escondido, Hyperion, and Point Loma for a much longer period of time than the other WTPs, we ran the above analyses two ways: all WTPs together from August – November 2020, and Escondido, Hyperion, and Point Loma samples for the full year separately.

We used HUMAnN3 (Beghini et al., 2021) with default settings to assign functional gene pathway annotations to reads using the UniRef90 (Suzek et al., 2015) and Metacyc (Caspi et al., 2020) databases. We also used RGI (the Resistance Gene Identifier) and the CARD and WildCARD databases (Alcock et al., 2020) to assign predicted antimicrobial resistance ontology identifiers (AROs) to the reads, then normalized all pathway abundances and AMR gene identities to transcripts per million (TPM). We compared microbial abundances, pathway abundances, and AMR gene abundances at greater than 0.01% relative abundance and present in 50% of samples between WTPs with ANCOM2.1 using sample collection month as an adjustment for covariates and sequencing batch as a random effect in the ANCOM models. We then plotted \log_{10} transformed counts of significantly differentially abundant viruses, bacterial genera, and AMR genes on a heatmap allowing the taxa to cluster with the Ward D2 algorithm with the R package "pheatmap" (Kolde, 2019). We used MaAsLin2 (Mallick et al., 2021) for longitudinal analyses of the above-mentioned variables, and included WTP and sequencing batch as random effects in the models, and we adjusted ANCOM and MaAsLin2

statistical tests for multiple comparisons with the Benjamini-Hochberg correction. We report the linear model coefficient with time of MaAsLin2 analyses on each plot and refer to Zenodo (doi.org/10.5281/zenodo.6829029) (Rothman et al., 2022) for individual scatterplots.

2.4. Data and code availability

Representative analyses scripts and code are available at github.com/jasonarothman/wastewater_metatranscriptomics_socal_aug20_aug21 and raw sequencing files have been deposited at the NCBI Sequence Read Archive under accession numbers PRJNA729801 using the MIMS: metagenome/environmental, wastewater; v5.0 package for each BioSample. Data tables containing taxa abundances, HUMAnN3 pathway annotations, and RGI assigned predicted antimicrobial resistance ontology identities are available as a Dryad dataset (<https://doi.org/10.7280/D11Q30>) (Rothman et al., 2022)

3. Results

3.1. Library statistics and microbial sample composition

We obtained a total of 4336,566,730 quality-filtered, nonhuman, paired-end reads across 275 samples from eight WTPs (average: 15,769,334 reads per sample, range: 1039,430 – 88,651,858, supplemental file SF1). With Kraken2, we classified an average of 55.0% of our reads (range 8.7 – 83.5%), of which an average of 48.1% (range 7.0 – 83.0%) were bacterial, 0.2% were archaeal (range 0.02 – 3.8%), and 5.9% (range 0.03 – 38.3%) were viral (Fig. 2). We also tried to classify nonhuman eukaryotes in our samples, but identified eukaryotes comprised less than 0.001% of reads. Due to the low relative abundance of archaea and nonhuman eukaryotes, and known questionable classification accuracy, we chose to focus on bacteria and viruses for diversity analyses.

We detected transcripts from a total of 6449 bacterial and 6888 viral species across all samples, however due to the likelihood of the taxonomic classifier reporting spurious species, we removed species accounting for < 0.01% average relative abundance within each domain. This filtering left us with 935 bacterial and 134 viral species present, which we used for downstream analyses. We also tabulated 245 bacterial families present in the same fashion as above. Because we had an uneven longitudinal distribution of samples, we analyzed diversity, differential abundance, and longitudinal relationships in two ways: First, samples where we had all eight WTPs were analyzed together representing $N = 98$, covering the months of August – November 2020. Second, we analyzed samples from Escondido, Hyperion, and Point Loma WTPs, where we had an entire year of sampling ($N = 214$), covering August 2020 – August 2021.

3.2. Antimicrobial resistance transcription

We detected transcripts matching 2128 unique antibiotic resistance ontology identifiers (AROs) through use of RGI and the CARD database (Fig. 2, Dryad: <https://doi.org/10.7280/D11Q30>). Of AROs that were present in an average of one read or greater per sample within each WTP (843 unique AROs), 25% were present at all WTPs, while 17% were unique to only one WTP (Fig. S1). AMR alpha diversity between August – November 2020 significantly differed between WTPs ($H_{(7)} = 33.7$, $P < 0.001$), but not over time ($t = -0.3$, $P = 0.74$) and did not correlate with the population served by WTP ($\rho = -0.25$, $P = 0.27$). AMR beta diversity during this time only differed between WTPs ($P < 0.001$, $R^2 = 0.43$), and not by month ($P = 0.08$, $R^2 = 0.03$), an interaction of WTP and month ($P = 0.10$, $R^2 = 0.16$), or sequencing batch ($P = 0.05$, $R^2 = 0.02$), and slightly changed over time ($t = -2.3$, $P = 0.03$) (Fig. 3). Several AMR transcripts were differentially abundant between WTPs, and for easier discrimination between the categories, we separated them into ribosomal RNA mutations and non-rRNA AMR genes: 29 rRNA AMR mutants

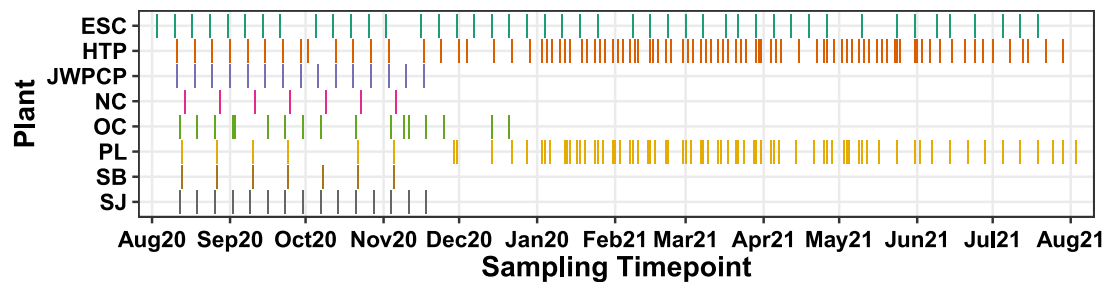


Fig. 1. Diagram indicating the date ranges of samples separated by wastewater treatment plan and month. Y-axis codes correspond to abbreviated WTP names: ESC = Escondido Hale Avenue Resource Recovery Facility, HTP = Hyperion Water Reclamation Plant, JWPCP = Joint Water Pollution Control Plant, NC = North City Water Reclamation Plant, OC = Orange County Reclamation Plant #1, PL = Point Loma Water Treatment Plant, SJ = San Jose Creek Water Reclamation Plant, and SB = South Bay Water Reclamation Plant.

($W > 88$, $P_{\text{adj}} < 0.05$) and 17 non-rRNA genes ($W > 140$, $P_{\text{adj}} < 0.05$) differed between WTPs (Fig. 3, supplemental file SF2).

When considering the entire year, AMR alpha diversity differed between WTPs ($H_{(2)} = 28.6$, $P < 0.001$), but not over time ($t = -0.6$, $P = 0.68$), and did not correlate with the population served by WTP ($\rho = 0.08$, $P = 0.27$). AMR beta diversity differed between WTPs ($P < 0.001$, $R^2 = 0.13$), month ($P < 0.001$, $R^2 = 0.16$), an interaction between WTP and month ($P = 0.007$, $R^2 = 0.12$), with significant sequencing batch effects ($P < 0.001$, $R^2 = 0.06$), and again, changed over time ($t = 3.3$, $P = 0.001$) (Fig. 3). We considered AMR transcripts from rRNA genes and non-ribosomal genes separately as above. For rRNA genes, we found that 26 positively and 13 negatively correlated with time ($P_{\text{adj}} < 0.05$), while 45 did not, and for non-ribosomal genes, 38 positively and 1 negatively changed over time ($P_{\text{adj}} < 0.05$), while 256 did not change significantly (Fig. 3, supplemental file SF3, Zenodo: <https://doi.org/10.5281/zenodo.6829029>).

3.3. Bacterial transcriptional ecology

We found that the top ten most proportionally abundant bacterial families represented an average of 58.6% (range 17.7 – 82.3%) of bacterial transcripts. These families (in descending average proportional abundance) were: Campylobacteraceae, Pseudomonadaceae, Enterobacteriaceae, Neisseriaceae, Moraxellaceae, Comamonadaceae, Burkholderiaceae, Aeromonadaceae, Weeksellaceae, and Methylobacteriaceae (Fig. 2).

As we only had samples from all WTPs from August – November 2020 we analyzed these separately, and found that bacterial transcript alpha diversity significantly differed between WTP ($H_{(7)} = 55.5$, $P < 0.001$), but not over time ($t = -1.3$, $P = 0.22$). Bacterial beta diversity was significantly different across WTPs ($P < 0.001$, $R^2 = 0.30$) and month ($P < 0.001$, $R^2 = 0.09$), with no interaction between WTP and month ($P = 0.13$, $R^2 = 0.16$), was affected by sequencing batch ($P < 0.001$, $R^2 = 0.03$), and changed over time ($t = -2.4$, $P = 0.02$). We also found that transcripts from 222/564 bacterial genera were significantly differentially abundant between WTPs during this time period ($W > 507$, $P_{\text{adj}} < 0.05$, Fig. 4, supplemental file SF2).

We had samples across the entire year for three WTPs, so we analyzed these independently, and found that alpha diversity was not different between WTPs ($H_{(2)} = 1.1$, $P = 0.59$), and did not differ over time ($t = 1.6$, $P = 0.12$). Beta diversity differed between WTP ($P < 0.001$, $R^2 = 0.07$), month ($P < 0.001$, $R^2 = 0.20$), and the interaction of WTP and month ($P = 0.002$, $R^2 = 0.11$) with significant sequencing batch effects ($P < 0.001$, $R^2 = 0.06$), and over time as a continuous variable ($t = 4.8$, $P < 0.001$). We tracked the transcription of bacterial genera across the year, and found that 172 genera increased, 63 genera decreased, and 295 did not change significantly over time (Fig. 4, supplemental file SF3, Zenodo: <https://doi.org/10.5281/zenodo.6829029>).

3.4. Viral ecology

We did not group viruses by family because of the dominance of Virgaviridae, and instead report summary statistics of the ten most proportionally abundant viral species as this provides more information. These viruses represented an average proportional viral abundance of 92.4% (range 33.1 – 99.5%; in descending average proportional abundance): Tomato brown rugose fruit virus, Cucumber green mottle mosaic virus, Pepper mild mottle virus, crAssphage, Tomato mosaic virus, Tropical soda apple mosaic virus, Tobacco mild green mosaic virus, Tomato mottle mosaic virus, Melon necrotic spot virus, and Pseudomonas virus PMBT3 (Fig. 2).

Over August – November 2020, viral alpha diversity differed between WTPs ($H_{(7)} = 35.1$, $P < 0.001$), but not over time ($t = -0.57$, $P = 0.58$). Beta diversity differed between WTPs ($P < 0.001$, $R^2 = 0.31$), by month ($P = 0.003$, $R^2 = 0.07$), but not by an interaction between WTP and month ($P = 0.69$, $R^2 = 0.13$), by sequencing batch ($P = 0.07$, $R^2 = 0.02$), or over time ($t = -1.6$, $P = 0.11$). During this time period, only 11 viruses were differentially abundant between WTPs (Fig. 5, supplemental file SF2).

The full-year samples showed significantly different alpha diversity between WTP ($H_{(2)} = 55.4$, $P < 0.001$) but not over time ($t = 0.11$, $P = 0.91$). Long-term beta diversity differed between WTPs ($P = 0.005$, $R^2 = 0.03$), month ($P = 0.001$, $R^2 = 0.11$), with no interaction between WTP and month ($P = 0.34$, $R^2 = 0.09$), with significant sequencing batch effects ($P = 0.001$, $R^2 = 0.11$), and changed significantly over time ($t = 4.3$, $P < 0.001$). When considering the proportional abundance of individual virus species over the year, 22 viruses increased 16 decreased, and 102 did not change over time (Fig. 5, supplemental file SF3, <https://doi.org/10.5281/zenodo.6829029>).

3.5. Metabolic pathway transcription

Across samples that successfully processed through HUMAnN3 ($N = 252$), we detected transcripts that mapped to 474 Metacyc metabolic pathways (Dryad: <https://doi.org/10.7280/D11Q30>). Most commonly, we found transcriptional activity from pathways such as nucleotide biosynthesis, ubiquitination, amino acid biosynthesis, and central carbon metabolism, while we also detected rarer pathways involved in the degradation of xenobiotics including toluene, atrazine, nitrobenzoate, and octane.

Metabolic transcript alpha diversity was not significantly different across WTPs from August – November 2020 ($H_{(7)} = 4.8$, $P = 0.68$) and did not change over time ($t = 1.2$, $P = 0.222$). Likewise, metabolic transcript beta diversity during this period was not different between WTPs ($P = 0.18$, $R^2 = 0.09$), but slightly differed between months ($P = 0.003$, $R^2 = 0.07$) with an interaction between month and WTP ($P = 0.03$, $R^2 = 0.24$) (Fig. 6), with significant sequencing batch effects ($P < 0.001$, $R^2 = 0.06$), but did not change over time ($t = 0.9$, $P = 0.38$).

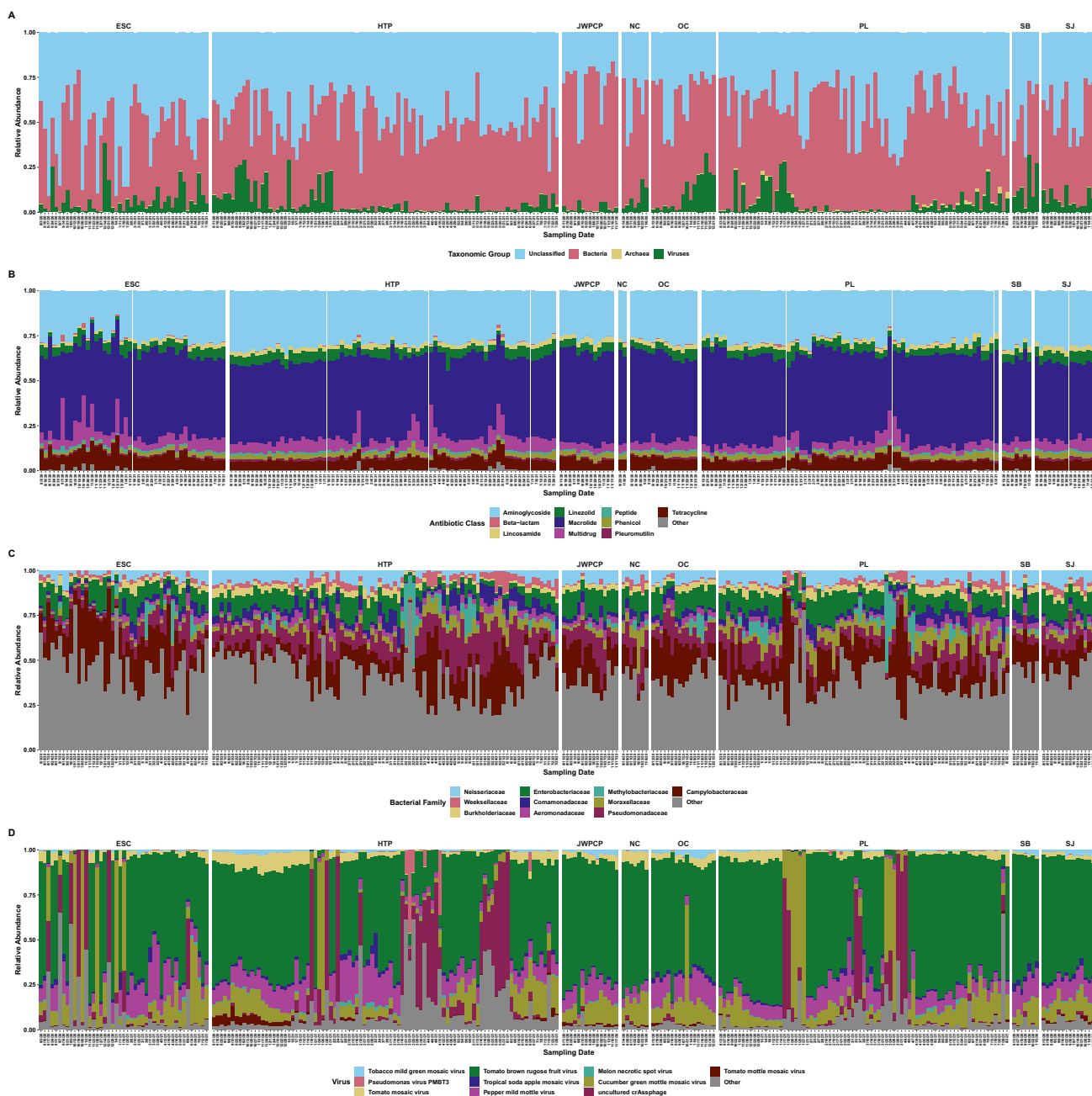


Fig. 2. Stacked bar plots showing the relative abundances of RNA reads mapping to A) unclassified taxonomic ranks, bacteria, viruses, and archaea; B) AMR genes separated by the ten most abundant antibiotic classes each gene confers resistance to plus all others; C) ten most abundant bacterial families plus all others; D) ten most abundant viral species plus all others. All plots are faceted by WTP and labeled with sampling date.

There were no differentially-expressed metabolic pathways across WTPs during this time period.

Across the full year, transcript alpha diversity differed between WTPs ($H_{(2)} = 14.4, P < 0.001$), but not over time ($t = 1.3, P = 0.19$). Beta diversity slightly differed between WTPs ($P = 0.008, R^2 = 0.02$), month ($P < 0.001, R^2 = 0.17$), with an interaction between WTP and month ($P = 0.005, R^2 = 0.12$) with significant sequencing batch effects ($P = 0.002, R^2 = 0.04$), and did not change longitudinally ($t = 0.8, P = 0.41$). The transcription of few metabolic pathways had a significant association with time, as only 12 were positively, and one was negatively correlated, out of 205 pathways total (Fig. 6, supplemental file SF3, Zenodo: <https://doi.org/10.5281/zenodo.6829029>).

4. Discussion

Composite wastewater samples from Southern California over the year contained RNA transcripts derived from a wide diversity of microorganisms. To the best of our knowledge, our study representing a sewer shed of 16 million people is the most complete metatranscriptomic characterization of a large metropolitan region's wastewater to date. Most notably, we show evidence of actively transcribed antimicrobial resistance (AMR) genes that encode resistance to a variety of commonly-administered antimicrobial drugs including macrolides, aminoglycosides, tetracycline and other AMR classes (Alcock et al., 2020). Likewise, we also show that bacterial transcription and RNA viral diversity differed between wastewater treatment plants (WTPs), and that sequencing wastewater RNA can be a useful tool for

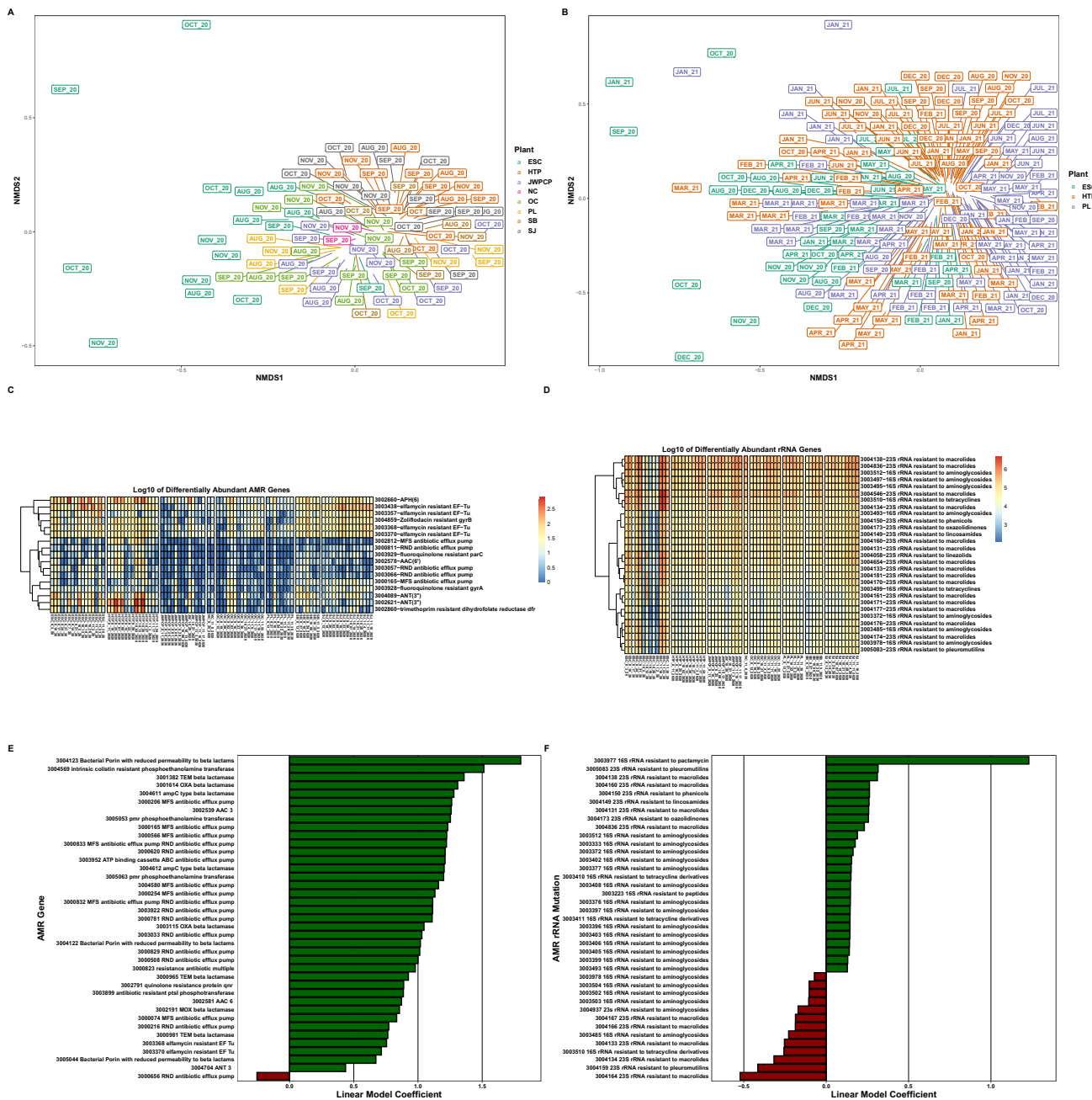


Fig. 3. Nonmetric multidimensional scaling ordination of Bray-Curtis dissimilarities of AMR genes at greater than 0.01% relative abundance across A) all WTPs August – November 2020, (NMDS stress = 0.08) and B) ESC, HTP, and PL across August 2020 – August 2021 (NMDS stress = 0.13). C) Heatmaps of the log₁₀-transformed counts of differentially abundant non-rRNA AMR genes across all WTPs August – November 2020, and D) rRNA gene mutations conferring resistance to antimicrobials. Hierarchical clustering of genes in each heatmap is through the Ward D2 algorithm. E) Bar plots indicating the non-rRNA AMR genes across ESC, HTP, and PL that changed over time and F) AMR rRNA gene mutations. X-axes denote the linear model coefficient of each gene’s relationship to time.

wastewater-based epidemiology (WBE) (Brumfield et al., 2022; Crits-Christoph et al., 2021; de Nies et al., 2021; Rothman et al., 2021; Xagorarakis and O’Brien, 2020). Finally, we examined the total RNA pool and described metabolic pathway transcription to show that wastewater metabolism is largely consistent across WTPs and over time, but that there are slight signatures of geographical location (Gulino et al., 2020). Likewise, we report that approximately 45% of our reads were unassigned by the Kraken2 database, which is not unusual for metagenomes even from well-studied environments and most likely represents microbial “dark matter,” along with sequencing artifacts (Hurwitz et al., 2016; Zund et al., 2022). The unknowns may also arise due to our Kraken2 database not including eukaryotes, however, we subsequently

attempted to classify eukaryotes using a different method (Lind and Pollard, 2021), but only classified an additional < 0.001% of the reads. Our results suggest that RNA sequencing is a viable tool to understand the complex matrix that wastewater represents and is useful in assaying the microbes associated with large populations.

4.1. Antimicrobial resistance transcription across southern California wastewater

Wastewater is known to harbor an array of AMR genes, and several studies have sequenced and/or quantified many of these genes in wastewater (de Nies et al., 2021; Ju et al., 2019; Raza et al., 2022; Yin

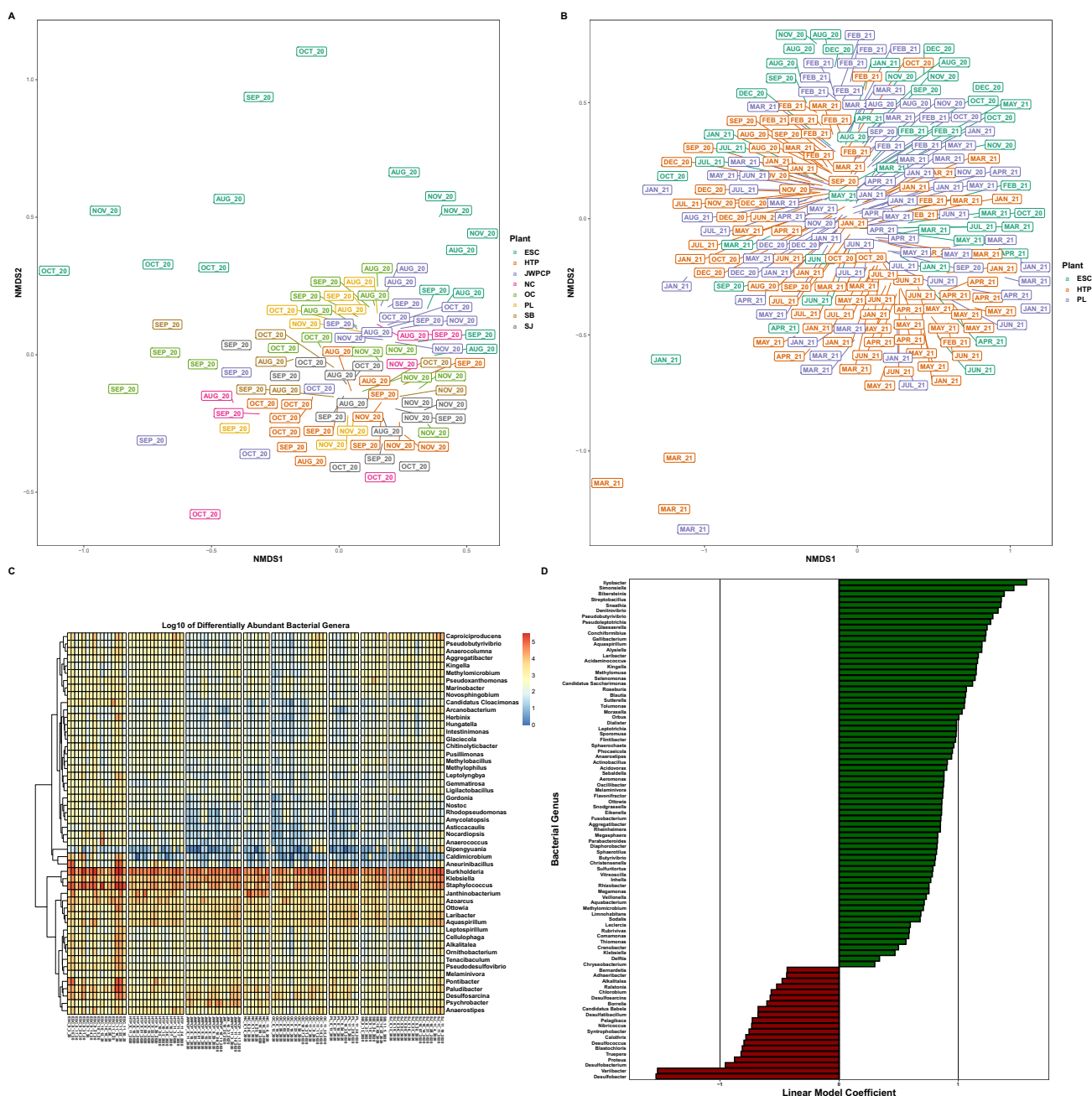


Fig. 4. Nonmetric multidimensional scaling ordination of Bray-Curtis dissimilarities of bacterial species at greater than 0.01% relative abundance across A) all WTPs August – November 2020 (NMDS stress = 0.11), and B) ESC, HTP, and PL across August 2020 – August 2021 (NMDS stress = 0.16). C) Heatmap of the log₁₀-transformed counts of differentially abundant bacterial genera at greater than 0.1% relative abundance across all WTPs August – November 2020. D) Bar plots indicating the bacterial genera across ESC, HTP, and PL that changed over time (only genera with a P_{adj} < 0.001 shown). X-axes denote the linear model coefficient of each genus’s relationship to time.

et al., 2021). Our study differs in that we demonstrate transcriptional activity through RNA-sequencing, rather than the genomic potential of the sampled organisms. We found a wide diversity of transcribed AMR genes that are commonly found in AMR infections across all WTPs, including components of the multidrug efflux pumps *adeFGH* (Coyné et al., 2010) and its repressor *acrS* (Hirakawa et al., 2008), the gene *tetQ* (Nikolich et al., 1992), which encodes a ribosomal protection protein against tetracycline, *Staphylococcus aureus*’s multidrug efflux protein *lmrS* (Floyd et al., 2010), genes in the aminoglycoside resistance series *aadA* and *aph(3’')* (Ramirez and Tolmasky, 2010), and several variants of the glycopeptide resistance gene *vanR* (Courvalin, 2006). Many of these transcripts have been previously detected in WTPs, or in animals

that resided in wastewater (Brumfield et al., 2022; Marcelino et al., 2019). Because we did not deplete rRNAs during library preparations, most of our bacterial transcripts were ribosomal RNAs. We detected rRNA mutations that confer macrolide resistance in the medically important taxa *Neisseria*, *Campylobacter*, *Salmonella*, *Helicobacter*, *Staphylococcus*, *Streptococcus*, *Klebsiella*, and many others. These genera (and subsequent antimicrobial-resistant rRNAs) were ubiquitous in our samples and are often found in wastewater (Jankowski et al., 2022; Joseph et al., 2019; Ju et al., 2019). Our results indicate transcriptional evidence of widespread AMR activity, and we posit that this AMR presence is likely to be found in other wastewater catchments making metatranscriptomics useful for tracking AMR across wide areas (de Nies

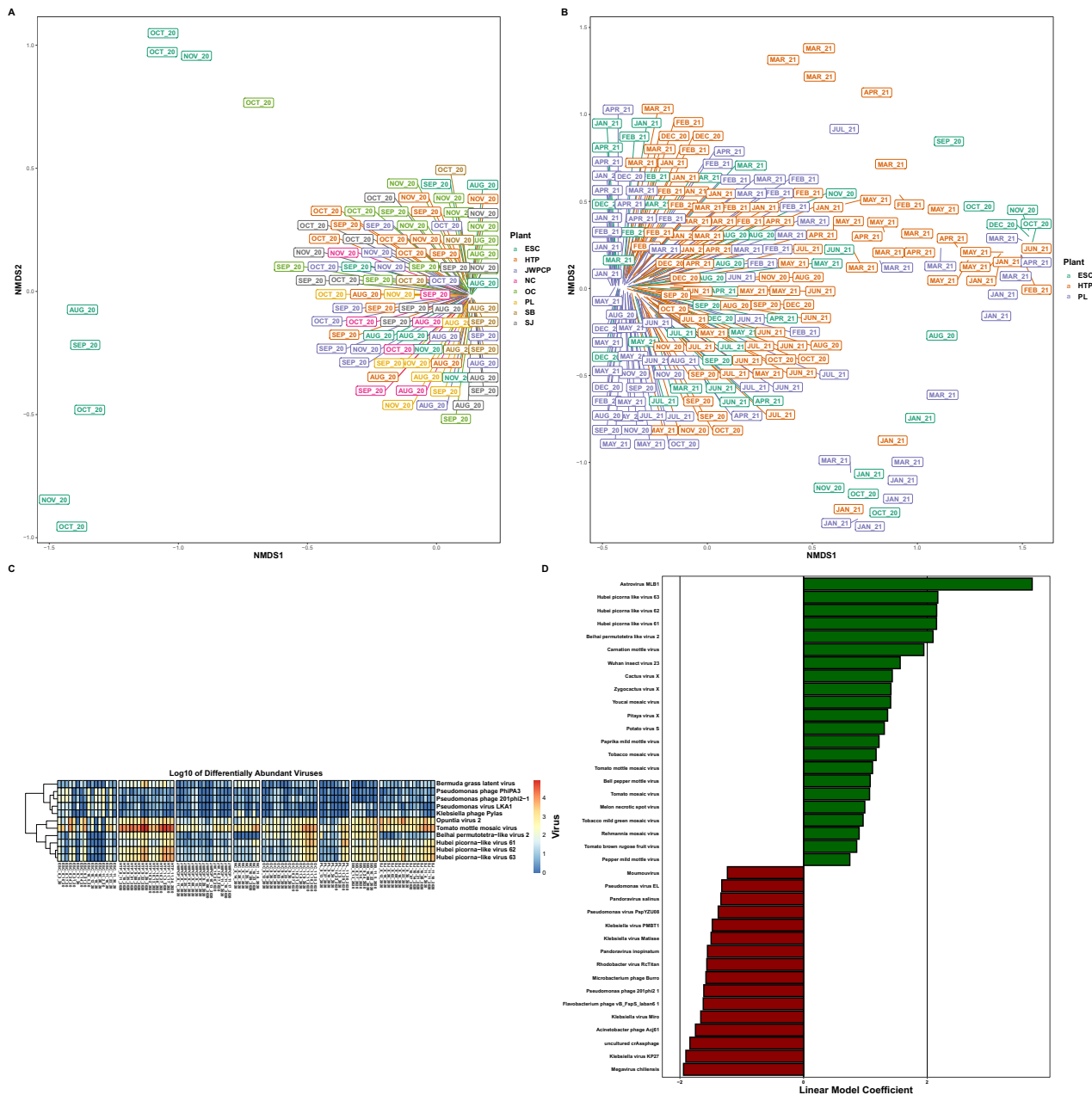


Fig. 5. Nonmetric multidimensional scaling ordination of Bray-Curtis dissimilarities of viruses at greater than 0.01% relative abundance across A) all WTPs August – November 2020 (NMDS stress = 0.10), and B) ESC, HTP, and PL across August 2020 – August 2021 (NMDS stress = 0.08). C) Heatmap of the log₁₀-transformed counts of differentially abundant viruses across all WTPs August – November 2020. D) Bar plots indicating the viruses across ESC, HTP, and PL that changed over time. X-axes denote the linear model coefficient of each virus’s relationship to time.

et al., 2021). The diversity of AMR genes in our samples differed between WTPs, and there were a few AMR genes differentially abundant between WTPs – mostly mutant rRNAs. This finding supports studies that show geographic differences between AMR (Raza et al., 2022; Yin et al., 2021), but there are likely other factors impacting the diversity of AMR, such as disease load in the served populations. Interestingly, we noticed a general increase over time in the proportional abundance of several transcripts from the major facilitator superfamily (MFS) and resistance-nodulation-cell division (RND) antibiotic efflux pumps - which are often implicated in multidrug resistance (Li and Nikaido, 2009) - along with beta-lactamases, and aminoglycoside/macrolide resistant rRNAs (Alcock et al., 2020). These data support studies showing an increase in antibiotic resistance (Ju et al., 2019) and the prevalence of AMR genes, but may also be impacted by seasonal changes

in the waste stream (Yang et al., 2013). While our results are suggestive of AMR becoming more prevalent, our results are likely not truly quantitative, especially given the technical challenges of sequencing wastewater (Karthikeyan et al., 2022). Similarly, our data reflect frequent detection of AMR genes in organisms that have been identified as “urgent” or “serious” by the Centers for Disease Control (CDC), although the release of some 2020 and 2021 data is delayed, so we are unable to truly link AMR infections to AMR gene prevalence (CDC, 2022; World Health Organization, 2021).

4.2. Viral ecology of southern California wastewater

Plant-infecting tobamoviruses dominated the viromes of our samples regardless of source or time of year (Bačnik et al., 2020; Brumfield et al.,

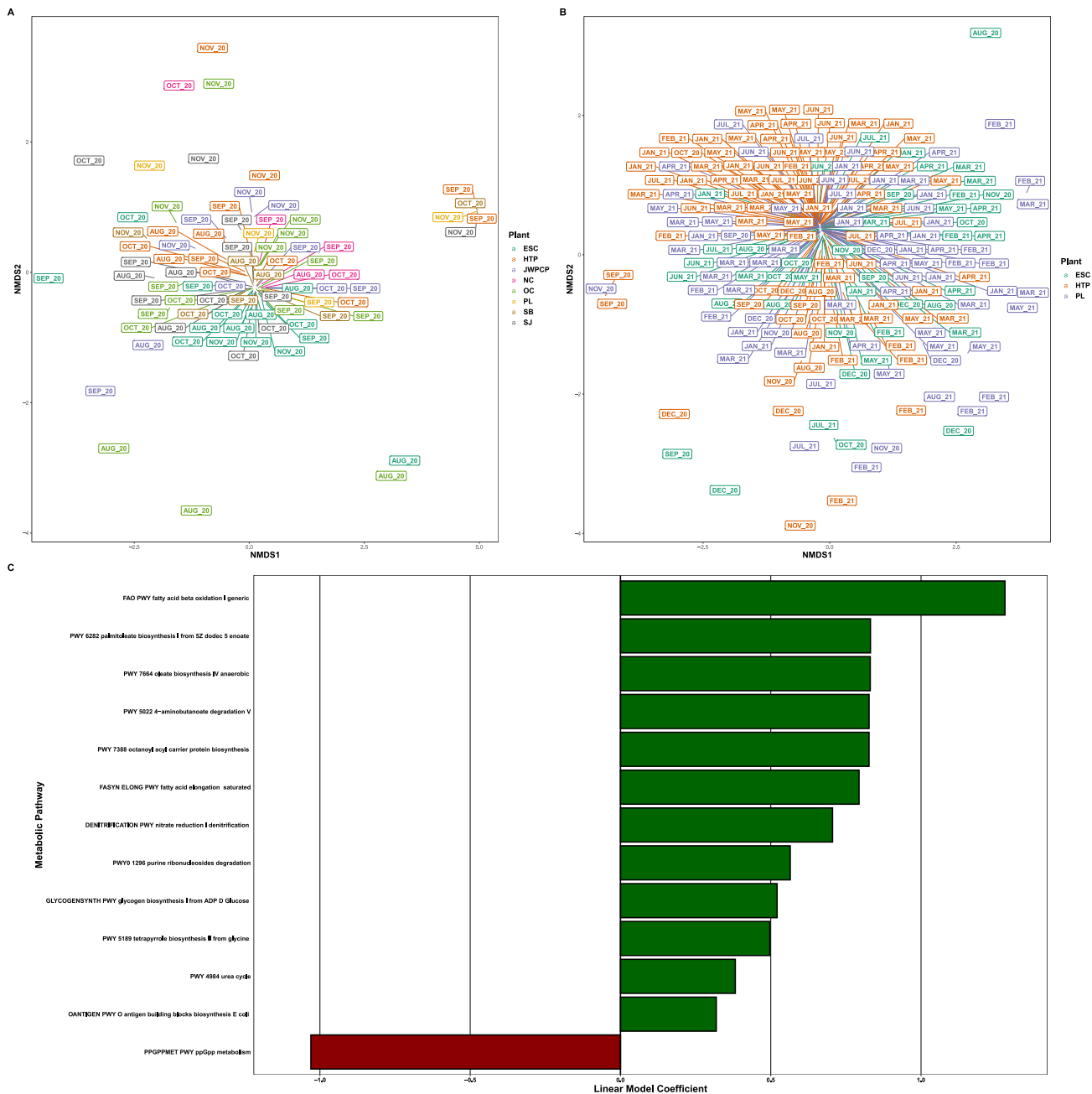


Fig. 6. Nonmetric multidimensional scaling ordination of Bray-Curtis dissimilarities of metabolic pathway transcripts per million across A) all WTPs August – November 2020 (NMDS stress = 0.07), and B) ESC, HTP, and PL across August 2020 – August 2021 (NMDS stress = 0.08). C) Bar plots indicating the metabolic pathway at greater than 0.01% relative abundance across ESC, HTP, and PL that changed over time. X-axes denote the linear model coefficient of each metabolic pathway’s relationship to time.

2022; Cantalupo et al., 2011; Crits-Christoph et al., 2021; Rothman and Whiteson, 2022), although we also found substantial numbers of reads mapping to arthropod-infecting viruses (i.e. Hubei picorna-like viruses, Wuhan insect viruses, and Flock House Virus), and phages including crAssphage and assorted bacteriophages. While most known phages have DNA genomes, previous studies have identified phages in wastewater RNA (Crits-Christoph et al., 2021; Wilder et al., 2021). We may be detecting novel RNA viruses, or transcription of either DNA or RNA based phage genomes, or there may be some DNA contamination from incomplete DNase treatment during RNA extraction confounding our results. Viral diversity differed when tested across all WTPs and over the full year, supporting studies that suggest geographical signatures of viruses in wastewater, and may be due to differences in human diet and

viral excretion, along with disease dynamics in bacteria and/or eukaryotic hosts (Bibby and Peccia, 2013; Brumfield et al., 2022; Gulino et al., 2020). Likewise, several viruses were differentially abundant over time, which may be due to underlying infection trends or due to unknown seasonality effects (Brinkman et al., 2017; Kazama et al., 2016). While overall viral diversity was different between WTPs and changed over time, highly abundant viruses tended to be present in most samples, which may afford new targets in establishing microbial water quality or the detection of sewage pollution (Cao et al., 2015; Jiang et al., 2022; Kitajima et al., 2018). Similarly, we detected several human-infecting viruses (i.e. Norwalk Virus in 78% of samples and SARS-CoV-2 in 62% of samples) which provides support for WBE efforts (Crits-Christoph et al., 2021; Nemudryi et al., 2020; Rothman et al., 2021, 2020);

Xagorarakis and O'Brien, 2020), but due to relatively few counts of these viruses of interest, we suggest that RNA sequencing of wastewater should be used in conjunction with targeted and quantificational approaches to assist in passively monitoring diseases across large populations.

4.3. Bacterial ecology and metabolic pathways in southern California wastewater

Similar to other studies, we detected transcripts from bacterial species in wastewater - mostly in the form of rRNA reads (de Nies et al., 2021; Joseph et al., 2019). Human pathogens were broadly represented in our data, including ESKAPE bacteria (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp.), *Campylobacter jejuni*, *Salmonella* spp., *Helicobacter pylori*, *Haemophilus* spp., sexually transmitted infectious (STIs) agents along with taxa associated with the human fecal microbiome previously found in wastewater such as *Bacteroides*, *Escherichia*, *Faecalibacterium*, and *Roseburia* spp (Liu et al., 2015; McLellan et al., 2010; Newton et al., 2015; Newton and McClary, 2019). As wastewater is a matrix and contains bacterial taxa from a variety of sewer-associated or environmental species, we also report finding appreciable relative abundances of transcripts from genera often found in wastewater such as *Arcobacter*, *Aeromonas*, and *Acidovorax* (Liu et al., 2015; McLellan et al., 2010). Much as with viruses, the bacterial profiles of WTPs were different, although many species were ubiquitous throughout the samples (Wu et al., 2019). There were also noticeable changes in the relative proportional transcript abundance over time, with many bacterial genera displaying a bimodal periodicity: Higher transcript abundance during Winter and Summer, and generally higher as time proceeded from August 2020 to August 2021. Other work has shown a distinct seasonality to the wastewater microbial community (Peces et al., 2022) - and our data supports this as well - although there are many other factors that can affect wastewater communities, such as pH, flux, dissolved oxygen, and detergents (Wu et al., 2019). Non-ribosomal bacterial metabolism was apparent in our data with transcripts mapping to widely-conserved pathways such as nucleotide and amino acid biosynthesis and ubiquitination, with no pathways differing between WTPs or over time (Caspi et al., 2020). Likewise, we recognize that our RNA extraction methods were harsh, and surely resulted in nucleic acid degradation, which likely affects the accuracy of our results (Schuierer et al., 2017). Similarly, our results may be confounded by the difficulty in sequencing wastewater samples, as wastewater-derived RNA is often fragmented, low-quality, low concentration, and contains PCR inhibitors (Karthikeyan et al., 2022). Collectively, our results suggest that sequencing bacterial species and their constituent metabolic pathways common to wastewater may be useful for monitoring disease through WBE, and that novel targets to assay microbial water quality may be possible.

5. Conclusion

In our opinion, this large-scale longitudinal dataset represents an unprecedented metatranscriptomic characterization of wastewater across a large population and region. We detected a wide diversity of transcribed AMR genes, suggesting that RNA sequencing is a powerful tool for WBE and may be useful in monitoring the spread and intensity of AMR. Within our study, we sequenced the viromes of a large portion of Southern California's wastewater catchment area and show that plant-infecting viruses dominate the RNA viral fraction, which may have additional uses in detecting agricultural disease outbreaks. Similarly, we detected numerous human pathogens and observed changes in the relative proportions of these taxa, lending more credence to WBE as a vital component to public health and microbial water quality assays. We suggest that future transcriptomic studies target disease-causing taxa in wastewater to understand and refine WBE and its usefulness to human

health more deeply.

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 and code are publicly available and links are included in the manuscript.

Acknowledgments

We thank the staff of the City of Escondido Hale Avenue Resource Recovery Facility, the City of Los Angeles Sanitation and Environment, Los Angeles County Sanitation District, Orange County Sanitation District, and City of San Diego Public Utilities for collecting influent.

This research was supported by the University of California Office of the President Research Grants Program Office (award numbers R01RG3732 and R00RG2814) awarded to JAR and KLW, and a Hewitt Foundation for Biomedical Research postdoctoral fellowship to JAR. This work was made possible, in part, through access to the Genomics High Throughput Facility Shared Resource of the Cancer Center Support Grant (P30CA-062203) at the University of California, Irvine, NIH shared instrumentation grants 1S10RR025496-01, 1S10OD010794-01, and 1S10OD021718-01, and access to computing resources from the UCI High Performance Cloud Computing Center.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2022.119421.

References

- Achak, M., Alaoui Bakri, S., Chhiti, Y., M'hamdi Alaoui, F.E., Barka, N., Boumya, W., 2021. SARS-CoV-2 in hospital wastewater during outbreak of COVID-19: a review on detection, survival and disinfection technologies. *Sci. Total Environ.* 761, 143192 <https://doi.org/10.1016/j.scitotenv.2020.143192>.
- Alcock, B.P., Raphenya, A.R., Lau, T.T.Y., Tsang, K.K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.-L.V., Cheng, A.A., Liu, S., Min, S.Y., Miroshnichenko, A., Tran, H.-K., Werfalli, R.E., Nasir, J.A., Oloni, M., Speicher, D.J., Florescu, A., Singh, B., Faltyn, M., Hernandez-Koutoucheva, A., Sharma, A.N., Bordeleau, E., Pawlowski, A.C., Zubyk, H.L., Dooley, D., Griffiths, E., Maguire, F., Winsor, G.L., Beiko, R.G., Brinkman, F.S.L., Hsiao, W.W.L., Domselaar, G.V., McArthur, A.G., 2020. CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 48, D517–D525. <https://doi.org/10.1093/nar/gkz935>.
- Amoah, I.D., Kumari, S., Bux, F., 2020. Coronaviruses in wastewater processes: source, fate and potential risks. *Environ. Int.* 143, 105962 <https://doi.org/10.1016/j.envint.2020.105962>.
- Bačnik, K., Kutnjak, D., Pecman, A., Mehle, N., Tušek Žnidarič, M., Gutiérrez Aguirre, I., Ravnikar, M., 2020. Viromics and infectivity analysis reveal the release of infective plant viruses from wastewater into the environment. *Water Res.* 177, 115628 <https://doi.org/10.1016/j.watres.2020.115628>.
- Beghini, F., McIver, L.J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A.M., Valles-Colomer, M., Weingart, G., Zhang, Y., Zolfo, M., Huttenhower, C., Franzosa, E.A., Segata, N., 2021. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife* 10. <https://doi.org/10.7554/eLife.65088>.
- Bibby, K., Peccia, J., 2013. Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environ. Sci. Technol.* 47, 1945–1951. <https://doi.org/10.1021/ce305181x>.
- Bivins, A., North, D., Ahmad, A., Ahmed, W., Alm, E., Been, F., Bhattacharya, P., Bijlsma, L., Boehm, A.B., Brown, J., Buttiglieri, G., Calabro, V., Carducci, A., Castiglioni, S., Cetecioglu Guro, Z., Chakraborty, S., Costa, F., Curcio, S., De Los Reyes, F.L., Delgado Vela, J., Farkas, K., Fernandez-Casi, X., Gerba, C., Gerrity, D., Girones, R., Gonzalez, R., Haramoto, E., Harris, A., Holden, P.A., Islam, M.T., Jones, D.L., Kasprzyk-Hordern, B., Kitajima, M., Kotlarz, N., Kumar, M., Kuroda, K., La Rosa, G., Malpei, F., Mautus, M., McLellan, S.L., Medema, G., Meschke, J.S., Mueller, J., Newton, R.J., Nilsson, D., Noble, R.T., Van Nuijs, A., Peccia, J.,

- Perkins, T.A., Pickering, A.J., Rose, J., Sanchez, G., Smith, A., Stadler, L., Stauber, C., Thomas, K., Van Der Voorn, T., Wigginton, K., Zhu, K., Bibby, K., 2020. Wastewater-based epidemiology: global collaboration to maximize contributions in the fight against COVID-19. *Environ. Sci. Technol.* <https://doi.org/10.1021/acs.est.0c02388>.
- Brinkman, N.E., Fout, G.S., Keely, S.P., 2017. Retrospective surveillance of wastewater to examine seasonal dynamics of Enterovirus infections. *mSphere* 2. <https://doi.org/10.1128/mSphere.00099-17>.
- Brumfield, K.D., Leddy, M., Usmani, M., Cotruvo, J.A., Tien, C.-T., Dorsey, S., Graubics, K., Fanelli, B., Zhou, I., Registe, N., Dadlani, M., Wimalarante, M., Jinasena, D., Abayagunawardena, R., Withanachchi, C., Huq, A., Jutla, A., Colwell, R.R., 2022. Microbiome analysis for wastewater surveillance during COVID-19. *MBio*, e0059122. <https://doi.org/10.1128/mbio.00591-22>.
- Bushnell, B., 2014. BBTools software package.
- Cantalupo, P.G., Calgua, B., Zhao, G., Hundesa, A., Wier, A.D., Katz, J.P., Grabe, M., Hendrix, R.W., Girones, R., Wang, D., Pipas, J.M., 2011. Raw sewage harbors diverse viral populations. *MBio* 2. <https://doi.org/10.1128/mbio.00180-11>.
- Cao, Y., Raith, M.R., Griffith, J.F., 2015. Droplet digital PCR for simultaneous quantification of general and human-associated fecal indicators for water quality assessment. *Water Res.* 70, 337–349. <https://doi.org/10.1016/j.watres.2014.12.008>.
- Caspi, R., Billington, R., Keseler, I.M., Kothari, A., Krummenacker, M., Midford, P.E., Ong, W.K., Paley, S., Subhraveti, P., Karp, P.D., 2020. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Res.* 48, D445–D453. <https://doi.org/10.1093/nar/gkz862>.
- CDC, 2022. COVID-19: U.S. Impact On Antimicrobial resistance, Special Report 2022. <https://doi.org/10.15620/cdc:117915>.
- Courvalin, P., 2006. Vancomycin resistance in gram-positive cocci. *Clin. Infect. Dis.* 42 (1), S25–S34. <https://doi.org/10.1086/491711>. Suppl.
- Coyne, S., Rosenfeld, N., Lambert, T., Courvalin, P., Périchon, B., 2010. Overexpression of resistance-nodulation-cell division pump AdefGH confers multidrug resistance in *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* 54, 4389–4393. <https://doi.org/10.1128/AAC.00155-10>.
- Crits-Christoph, A., Kantor, R.S., Olm, M.R., Whitney, O.N., Al-Shayeb, B., Lou, Y.C., Flamholz, A., Kennedy, L.C., Greenwald, H., Hinkle, A., Hetzel, J., Spitzer, S., Koble, J., Tan, A., Hyde, F., Schroth, G., Kuersten, S., Banfield, J.F., Nelson, K.L., 2021. Genome sequencing of sewage detects regionally prevalent SARS-CoV-2 Variants. *MBio* 12. <https://doi.org/10.1128/mbio.02703-20>.
- de Nies, L., Busi, S.B., Kunath, B.J., May, P., Wilmes, P., 2021. Mobilome-driven segregation of the resistome in biological wastewater treatment. *bioRxiv*. <https://doi.org/10.1101/2021.11.15.468621>.
- Edwards, R.A., Vega, A.A., Norman, H.M., Ohaeri, M., Levi, K., Dinsdale, E.A., 2019. Global phylogeography and ancient evolution of the widespread human gut virus crAsphage. *Nat. Microbiol.* 4, 1727–1736. <https://doi.org/10.1038/s41564-019-0494-6>.
- Farkas, K., Adriaenssens, E.M., Walker, D.I., McDonald, J.E., Malham, S.K., Jones, D.L., 2019. Critical evaluation of crAsphage as a molecular marker for human-derived wastewater contamination in the aquatic environment. *Food Environ. Virol.* 11, 113–119. <https://doi.org/10.1007/s12560-019-09369-1>.
- Faust, K., Lahti, L., Gonze, D., de Vos, W.M., Raes, J., 2015. Metagenomics meets time series analysis: unraveling microbial community dynamics. *Curr. Opin. Microbiol.* 25, 56–66. <https://doi.org/10.1016/j.mib.2015.04.004>.
- Floyd, J.L., Smith, K.P., Kumar, S.H., Floyd, J.T., Varela, M.F., 2010. LmrS is a multidrug efflux pump of the major facilitator superfamily from *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 54, 5406–5412. <https://doi.org/10.1128/AAC.00580-10>.
- Gulino, K., Rahman, J., Badri, M., Morton, J., Bonneau, R., Ghedin, E., 2020. Initial mapping of the New York City wastewater virome. *mSystems* 5. <https://doi.org/10.1128/mSystems.00876-19>.
- Hellmér, M., Paxéus, N., Magnius, L., Enache, L., Arnholm, B., Johansson, A., Bergström, T., Norder, H., 2014. Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks. *Appl. Environ. Microbiol.* 80, 6771–6781. <https://doi.org/10.1128/AEM.01981-14>.
- Hirakawa, H., Takumi-Kobayashi, A., Theisen, U., Hirata, T., Nishino, K., Yamaguchi, A., 2008. AcrS/EnvR represses expression of the acrAB multidrug efflux genes in *Escherichia coli*. *J. Bacteriol.* 190, 6276–6279. <https://doi.org/10.1128/JB.00190-08>.
- Hubeny, J., Korzeniewska, E., Buta-Hubeny, M., Zieliński, W., Rolbiecki, D., Harnisz, M., 2022. Characterization of carbapenem resistance in environmental samples and *Acinetobacter* spp. isolates from wastewater and river water in Poland. *Sci. Total Environ.* 822, 153437. <https://doi.org/10.1016/j.scitotenv.2022.153437>.
- Hurwitz, B.L., U'Ren, J.M., Youens-Clark, K., 2016. Computational prospecting the great viral unknown. *FEMS Microbiol. Lett.* 363. <https://doi.org/10.1093/femsle/fnw077>.
- Jankowski, P., Gan, J., Le, T., McKennitt, M., Garcia, A., Yanaç, K., Yuan, Q., Uyaguari-Diaz, M., 2022. Metagenomic community composition and resistome analysis in a full-scale cold climate wastewater treatment plant. *Environ. Microbiol.* 17, 3. <https://doi.org/10.1186/s40793-022-00398-1>.
- Jiang, S.C., Bischel, H.N., Goel, R., Rosso, D., Sherchan, S.P., Whiteson, K.L., Yan, T., Solo-Gabriele, H.M., 2022. Integrating virus monitoring strategies for safe non-potable water reuse. *Water (Basel)* 14, 1187. <https://doi.org/10.3390/w14081187>.
- Joseph, S.M., Battaglia, T., Maritz, J.M., Carlton, J.M., Blaser, M.J., 2019. Longitudinal comparison of bacterial diversity and antibiotic resistance genes in New York City sewage. *mSystems* 4. doi:10.1128/mSystems.00327-19.
- Ju, F., Beck, K., Yin, X., Maccagnan, A., McArdell, C.S., Singer, H.P., Johnson, D.R., Zhang, T., Bürgmann, H., 2019. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. *ISME J.* 13, 346–360. <https://doi.org/10.1038/s41396-018-0277-8>.
- Karkman, A., Do, T.T., Walsh, F., Virta, M.P.J., 2018. Antibiotic-resistance genes in waste water. *Trends Microbiol.* 26, 220–228. <https://doi.org/10.1016/j.tim.2017.09.005>.
- Karthikeyan, S., Levy, J.I., De Hoff, P., Humphrey, G., Birmingham, A., Jepsen, K., Farmer, S., Tubb, H.M., Valles, T., Tribelhorn, C.E., Tsai, R., Aigner, S., Sathe, S., Moshiri, N., Henson, B., Mark, A.M., Hakim, A., Baer, N.A., Ramesh, T., Belda-Ferre, P., Chacón, M., Cheung, W., Cresini, E.S., Eisner, E.R., Lastrella, A.L., Lawrence, E.S., Marotz, C.A., Ngo, T.T., Neustrander, T., Plascencia, A., Salido, R.A., Seaver, P., Smoot, E.W., McDonald, D., Neuhaard, R.M., Scioscia, A.L., Satterlund, A.M., Simmons, E.H., Abelman, D.B., Brenner, D., Bruner, J.C., Buckley, A., Ellison, M., Gattas, J., Gonias, S.L., Hale, M., Hawkins, F., Ikeda, L., Jhaveri, H., Johnson, T., Kellen, V., Kremer, B., Matthews, G., McLawhorn, R.W., Ouillet, P., Park, D., Pradenas, A., Reed, S., Riggs, L., Sanders, A., Sollenberger, B., Song, A., White, B., Winbush, T., Aceves, C.M., Anderson, Catelyn, Gangavarapu, K., Hufbauer, E., Kurzban, E., Lee, J., Matteson, N.L., Parker, R., Perkins, S.A., Robles-Kisikisaka, R., Schwab, M.A., Spencer, E., Wohl, S., Nicholson, L., McHardy, I.H., Dimmock, D.P., Hobbs, C.A., Bakhtar, O., Harding, A., Mendoza, A., Bolze, A., Kingsmore, S.F., Lee, W., Shah, S., McDonald, E., Yu, A.T., Zeller, M., Fisch, K.M., Longhurst, C., Maysent, P., Pride, D., Khosla, P.K., Laurent, L.C., Yeo, G.W., Andersen, K.G., Knight, R., 2022. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. *Nature* 609, 101–108. <https://doi.org/10.1038/s41586-022-05049-6>.
- Karthikeyan, S., Ronquillo, N., Belda-Ferre, P., Alvarado, D., Javidi, T., Longhurst, C.A., Knight, R., 2021. High-throughput wastewater SARS-CoV-2 detection enables forecasting of community infection dynamics in San Diego County. *mSystems* 6. <https://doi.org/10.1128/mSystems.00045-21>.
- Kazama, S., Masago, Y., Tohma, K., Souma, N., Imagawa, T., Suzuki, A., Liu, X., Saito, M., Oshitani, H., Omura, T., 2016. Temporal dynamics of norovirus determined through monitoring of municipal wastewater by pyrosequencing and virological surveillance of gastroenteritis cases. *Water Res.* 92, 244–253. <https://doi.org/10.1016/j.watres.2015.10.024>.
- Kitajima, M., Sassi, H.P., Torrey, J.R., 2018. Pepper mild mottle virus as a water quality indicator. *NPJ Clean Water* 1, 19. <https://doi.org/10.1038/s41545-018-0019-5>.
- Kolde, R., 2019. pheatmap: pretty Heatmaps.
- Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. {lmerTest} Package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. <https://doi.org/10.18637/jss.v082.i13>.
- Langford, B.J., So, M., Raybardhan, S., Leung, V., Westwood, D., MacFadden, D.R., Soucy, J.-P.R., Daneman, N., 2020. Bacterial co-infection and secondary infection in patients with COVID-19: a living rapid review and meta-analysis. *Clin. Microbiol. Infect.* 26, 1622–1629. <https://doi.org/10.1016/j.cmi.2020.07.016>.
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.
- Li, R., Zhu, L., Cui, L., Zhu, Y.-G., 2022. Viral diversity and potential environmental risk in microplastic at watershed scale: evidence from metagenomic analysis of plastsphere. *Environ. Int.* 161, 107146. <https://doi.org/10.1016/j.envint.2022.107146>.
- Li, X.-Z., Nikaïdo, H., 2009. Efflux-mediated drug resistance in bacteria: an update. *Drugs* 69, 1555–1623. <https://doi.org/10.2165/11317030-000000000-00000>.
- Lind, A.L., Pollard, K.S., 2021. Accurate and sensitive detection of microbial eukaryotes from whole metagenome shotgun sequencing. *Microbiome* 9, 58. <https://doi.org/10.1186/s40168-021-01015-y>.
- Liu, Y., Dong, Q., Shi, H., 2015. Distribution and population structure characteristics of microorganisms in urban sewage system. *Appl. Microbiol. Biotechnol.* 99, 7723–7734. <https://doi.org/10.1007/s00253-015-6661-7>.
- Lu, J., Breitwieser, F.P., Thielen, P., Salzberg, S.L., 2017. Bracken: estimating species abundance in metagenomics data. *PeerJ. Computer Science* 2017, e104. <https://doi.org/10.7717/peerj-cs.104>.
- Lu, L., Guest, J.S., Peters, C.A., Zhu, X., Rau, G.H., Ren, Z.J., 2018. Wastewater treatment for carbon capture and utilization. *Nat. Sustain.* 1, 750–758. <https://doi.org/10.1038/s41893-018-0187-9>.
- Mallick, H., Rahnavard, A., McIver, L.J., Ma, S., Zhang, Y., Nguyen, L.H., Tickle, T.L., Weingart, G., Ren, B., Schwager, E.H., Chatterjee, S., Thompson, K.N., Wilkinson, J. E., Subramanian, A., Lu, Y., Waldron, L., Paulson, J.N., Franzosa, E.A., Bravo, H.C., Huttenhower, C., 2021. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput. Biol.* 17, e1009442. <https://doi.org/10.1371/journal.pcbi.1009442>.
- Manor, Y., Handscher, R., Halmut, T., Neuman, M., Bobrov, A., Rudich, H., Vonsover, A., Shulman, L., Kew, O., Mendelson, E., 1999. Detection of poliovirus circulation by environmental surveillance in the absence of clinical cases in Israel and the Palestinian Authority. *J. Clin. Microbiol.* 37, 1670–1675. <https://doi.org/10.1128/jcm.37.6.1670-1675.1999>.
- Marcelino, V.R., Wille, M., Hurt, A.C., González-Acuña, D., Klaassen, M., Schlub, T.E., Eden, J.-S., Shi, M., Iredell, J.R., Sorrell, T.C., Holmes, E.C., 2019. Meta-transcriptomics reveals a diverse antibiotic resistance gene pool in avian microbiomes. *BMC Biol.* 17, 31. <https://doi.org/10.1186/s12915-019-0649-1>.
- Martínez-Puchol, S., Rusiñol, M., Fernández-Cassi, X., Timonedá, N., Itarte, M., Andrés, C., Antón, A., Abril, J.F., Girones, R., Bofill-Mas, S., 2020. Characterisation of the sewage virome: comparison of NGS tools and occurrence of significant pathogens. *Sci. Total Environ.* 713, 136604. <https://doi.org/10.1016/j.scitotenv.2020.136604>.

- McLellan, S.L., Huse, S.M., Mueller-Spitz, S.R., Andreishcheva, E.N., Sogin, M.L., 2010. Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environ. Microbiol.* 12, 378–392. <https://doi.org/10.1111/j.1462-2920.2009.02075.x>.
- Nemudryi, A., Nemudraia, A., Wiegand, T., Surya, K., Buyukyoruk, M., Cicha, C., Vanderwood, K.K., Wilkinson, R., Wiedenheft, B., 2020. Temporal detection and phylogenetic assessment of SARS-CoV-2 in municipal wastewater. *Cell Rep. Med.* 1, 100098 <https://doi.org/10.1016/j.crm.2020.100098>.
- Newton, R.J., McClary, J.S., 2019. The flux and impact of wastewater infrastructure microorganisms on human and ecosystem health. *Curr. Opin. Biotechnol.* 57, 145–150. <https://doi.org/10.1016/j.copbio.2019.03.015>.
- Newton, R.J., McLellan, S.L., Dila, D.K., Vineis, J.H., Morrison, H.G., Eren, A.M., Sogin, M.L., 2015. Sewage reflects the microbiomes of human populations. *MBio* 6, e02574. <https://doi.org/10.1128/mBio.02574-14>.
- Nikolich, M.P., Shoemaker, N.B., Salyers, A.A., 1992. A *Bacteroides* tetracycline resistance gene represents a new class of ribosome protection tetracycline resistance. *Antimicrob. Agents Chemother.* 36, 1005–1012. <https://doi.org/10.1128/AAC.36.5.1005>.
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2017. *vegan: community Ecology Package*.
- Peccia, J., Zulli, A., Brackney, D.E., Grubaugh, N.D., Kaplan, E.H., Casanovas-Massana, A., Ko, A.I., Malik, A.A., Wang, D., Wang, M., Warren, J.L., Weinberger, D. M., Arnold, W., Omer, S.B., 2020. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nat. Biotechnol.* 38, 1164–1167. <https://doi.org/10.1038/s41587-020-0684-z>.
- Peces, M., Dottorini, G., Nierychlo, M., Andersen, K.S., Dueholm, M.K.D., Nielsen, P.H., 2022. Microbial communities across activated sludge plants show recurring species-level seasonal patterns. *ISME commun.* 2, 1–11. <https://doi.org/10.1038/s43705-022-00098-4>.
- Pedersen, T.L., 2020. Patchwork: the composer of plots. R package version 1, 182.
- Ramirez, M.S., Tolmashy, M.E., 2010. Aminoglycoside modifying enzymes. *Drug Resist. Updat.* 13, 151–171. <https://doi.org/10.1016/j.drug.2010.08.003>.
- Raza, S., Shin, H., Hur, H.-G., Unno, T., 2022. Higher abundance of core antimicrobial resistant genes in effluent from wastewater treatment plants. *Water Res.* 208, 117882 <https://doi.org/10.1016/j.watres.2021.117882>.
- Rose, A.N., Baggs, J., Wolford, H., Neuhauser, M.M., Srinivasan, A., Gundlapalli, A.V., Reddy, S., Kompanyets, L., Pennington, A.F., Grigg, C., Kabbani, S., 2021. Trends in antibiotic use in united states hospitals during the Coronavirus Disease 2019 Pandemic. *Open Forum Infect. Dis.* 8, ofab236. <https://doi.org/10.1093/ofid/ofab236>.
- Rothman, J., Sagir, A., Chung, S.-A., Boyajian, N., Dinh, T., Kim, J., Oval, J., Sharavanan, V., York, C., Zimmer-Faust, A., Langlois, K., Steele, J., Griffith, J., Whiteson, K., 2022. Data for: longitudinal metatranscriptomic sequencing of Southern California wastewater representing 16 million people from August 2020-21 reveals widespread transcription of antibiotic resistance genes. doi:10.7280/D11 Q30.
- Rothman, J.A., Loveless, T.B., Griffith, M.L., Steele, J.A., Griffith, J.F., Whiteson, K.L., 2020. Metagenomics of wastewater influent from Southern California wastewater treatment facilities in the era of COVID-19. *Microbiol. Resour. Announc.* 9, 19–21. <https://doi.org/10.1128/mra.00907-20>.
- Rothman, J.A., Loveless, T.B., Karpia 3rd, J., Adams, E.D., Steele, J.A., Zimmer-Faust, A. G., Langlois, K., Wanless, D., Griffith, M., Mao, L., Chokry, J., Griffith, J.F., Whiteson, K.L., 2021. RNA viromics of Southern California wastewater and detection of SARS-CoV-2 single-nucleotide variants. *Appl. Environ. Microbiol.* 87, e0144821 <https://doi.org/10.1128/AEM.01448-21>.
- Rothman, J.A., Whiteson, K.L., 2022. Sequencing and variant detection of eight abundant plant-infecting tobamoviruses across Southern California wastewater. *Microbiol. Spectr.* <https://doi.org/10.1128/spectrum.03050-22> e03050-22.
- Schuerer, S., Carbone, W., Knehr, J., Petitjean, V., Fernandez, A., Sultan, M., Roma, G., 2017. A comprehensive assessment of RNA-seq protocols for degraded and low-quantity samples. *BMC Genom.* 18, 442. <https://doi.org/10.1186/s12864-017-3827-y>.
- Sims, N., Kasprzyk-Hordern, B., 2020. Future perspectives of wastewater-based epidemiology: monitoring infectious disease spread and resistance to the community level. *Environ. Int.* 139, 105689 <https://doi.org/10.1016/j.envint.2020.105689>.
- Slowikowski, K., 2018. ggprel: automatically position non-overlapping text labels with “ggplot2.” R package version 0. 8. 0.
- Suzek, B.E., Wang, Y., Huang, H., McGarvey, P.B., Wu, C.H., Consortium, UniProt, 2015. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 31, 926–932. <https://doi.org/10.1093/bioinformatics/btu739>.
- Symonds, E.M., Griffin, D.W., Breitbart, M., 2009. Eukaryotic viruses in wastewater samples from the United States. *Appl. Environ. Microbiol.* 75, 1402–1409. <https://doi.org/10.1128/AEM.01899-08>.
- Wickham, H., 2009. *ggplot2: Elegant graphics For Data Analysis*. Springer-Verlag, New York.
- Wilder, M.L., Middleton, F., Larsen, D.A., Du, Q., Fenty, A., Zeng, T., Insaft, T., Kilaru, P., Collins, M., Kmush, B., Green, H.C., 2021. Co-quantification of crAssphage increases confidence in wastewater-based epidemiology for SARS-CoV-2 in low prevalence areas. *Water Res.* X 11, 100100. <https://doi.org/10.1016/j.wroa.2021.100100>.
- Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20, 257. <https://doi.org/10.1186/s13059-019-1891-0>.
- World Health Organization, 2021. *Global Antimicrobial Resistance and Use Surveillance System (GLASS) report: 2021*. World Health Organization.
- Wu, F., Zhang, J., Xiao, A., Gu, X., Lee, W.L., Armas, F., Kauffman, K., Hanage, W., Matus, M., Ghaeli, N., Endo, N., Duvallet, C., Poyet, M., Moniz, K., Washburne, A.D., Erickson, T.B., Chai, P.R., Thompson, J., Alm, E.J., 2020. SARS-CoV-2 titers in wastewater are higher than expected from clinically confirmed cases. *mSystems* 5, e00614–e00620. <https://doi.org/10.1128/mSystems.00614-20>.
- Global Water Microbiome Consortium Wu, L., Ning, D., Zhang, B., Li, Y., Zhang, P., Shan, X., Zhang, Q., Brown, M.R., Li, Z., Van Nostrand, J.D., Ling, F., Xiao, N., Zhang, Y., Vierheilig, J., Wells, G.F., Yang, Y., Deng, Y., Tu, Q., Wang, A., Zhang, T., He, Z., Keller, J., Nielsen, P.H., Alvarez, P.J.J., Criddle, C.S., Wagner, M., Tiedje, J. M., He, Q., Curtis, T.P., Stahl, D.A., Alvarez-Cohen, L., Rittmann, B.E., Wen, X., Zhou, J., 2019. Global diversity and biogeography of bacterial communities in wastewater treatment plants. *Nat. Microbiol.* 4, 1183–1195. <https://doi.org/10.1038/s41564-019-0426-5>.
- Xagorarakis, I., O'Brien, E., 2020. Wastewater-based epidemiology for early detection of viral outbreaks. *Women in Water Quality* 75. https://doi.org/10.1007/978-3-030-17819-2_5.
- Yang, Y., Li, B., Ju, F., Zhang, T., 2013. Exploring variation of antibiotic resistance genes in activated sludge over a four-year period through a metagenomic approach. *Environ. Sci. Technol.* 47, 10197–10205. <https://doi.org/10.1021/es4017365>.
- Yin, X., Yang, Y., Deng, Y., Huang, Y., Li, L., Chan, L.Y.L., Zhang, T., 2021. An assessment of resistome and mobilome in wastewater treatment plants through temporal and spatial metagenomic analysis. *Water Res.* 209, 117885 <https://doi.org/10.1016/j.watres.2021.117885>.
- Zimmer-Faust, A.G., Steele, J.A., Xiong, X., Staley, C., Griffith, M., Sadowsky, M.J., Diaz, M., Griffith, J.F., 2021. A combined digital PCR and next generation DNA-sequencing based approach for tracking nearshore pollutant dynamics along the southwest United States/Mexico Border. *Front. Microbiol.* 12, 674214 <https://doi.org/10.3389/fmicb.2021.674214>.
- Zund, M., Dunham, S.J.B., Rothman, J.A., Whiteson, K.L., 2022. What lies beneath? Taking the plunge into the murky waters of phage biology. *mSystems*. Accepted manuscript.