1 Title: Longitudinal metatranscriptomic sequencing of Southern California wastewater

2 representing 16 million people from August 2020-21 reveals widespread transcription of

3 antibiotic resistance genes.

4

5 Jason A. Rothman[a#], Andrew Saghir[a], Seung-Ah Chung[b], Nicholas Boyajian[a], Thao Dinh[a],

6 Jinwoo Kim[a], Jordan Oval[a], Vivek Sharavanan[a], Courtney York[a], Amity G. Zimmer-Faust[c],

7 Kylie Langlois[c], Joshua A. Steele[c], John F. Griffith[c], Katrine L. Whiteson[a#]

8

9 [a] Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine,

10 CA, USA.

11 [b] Genomics High-Throughput Facility, Department of Biological Chemistry, University of

12 California, Irvine, Irvine, CA, USA.

13 [c] Southern California Coastal Water Research Project, Costa Mesa, CA, USA

14

15 # Co-corresponding authors: Jason A. Rothman, University of California, Irvine, Irvine, CA,

16 92697, (949) 824-3509, rothmanj@uci.edu. Katrine L. Whiteson, University of California,

17 Irvine, Irvine, CA, 92697, (949) 824-9032, katrine@uci.edu.

18   Abstract:

19   Municipal wastewater provides a representative sample of human fecal waste across a

20   catchment area and contains a wide diversity of microbes. Sequencing wastewater samples

21   provides information about human-associated and medically-important microbial populations,

22   and may be useful to assay disease prevalence and antimicrobial resistance (AMR).

23   Here, we present a study in which we used untargeted metatranscriptomic sequencing on

24   RNA extracted from 275 sewage influent samples obtained from eight wastewater treatment

25   plants (WTPs) representing approximately 16 million people in Southern California between

26   August 2020 – August 2021. We characterized bacterial and viral transcripts, assessed metabolic

27   pathway activity, and identified over 2,000 AMR genes/variants across all samples. Because we

28   did not deplete ribosomal RNA, we have a unique window into AMR carried as ribosomal

29   mutants. We show that AMR diversity varied between WTPs and that the relative abundance of

30   many individual AMR genes/variants increased over time and may be connected to antibiotic use

31   during the COVID-19 pandemic. Similarly, we detected transcripts mapping to human

32   pathogenic bacteria and viruses suggesting RNA sequencing is a powerful tool for wastewater-

33   based epidemiology and that there are geographical signatures to microbial transcription. We

34   captured the transcription of gene pathways common to bacterial cell processes, including central

35   carbon metabolism, nucleotide synthesis/salvage, and amino acid biosynthesis. We also posit that

36   due to the ubiquity of many viruses and bacteria in wastewater, new biological targets for

37   microbial water quality assessment can be developed.

38   To the best of our knowledge, our study provides the most complete longitudinal

39   metatranscriptomic analysis of a large population's wastewater to date and demonstrates our

40   ability to monitor the presence and activity of microbes in complex samples. By sequencing

41   RNA, we can track the relative abundance of expressed AMR genes/variants and metabolic

42   pathways, increasing our understanding of AMR activity across large human populations and

43   sewer sheds.

44

45   Keywords: Wastewater, antimicrobial resistance, metatranscriptomics, microbial ecology,

46   environmental microbiology.

47

48   1. Introduction:

49   Wastewater harbors a wide diversity of microorganisms and represents the collective

50   waste of human activity across a sewershed (Newton and McClary, 2019). Over 300 km^3 of

51   wastewater is produced globally, of which most is channeled into wastewater treatment plants

52   (WTPs) for biological and chemical processing (Lu et al., 2018). As a heterogenous mixture,

53   wastewater has been shown to contain microbial communities that vary depending on sampling

54   location, time of year, industry, agriculture, and the health of the served human population

55   (Cantalupo et al., 2011; Edwards et al., 2019; McLellan et al., 2010; Symonds et al., 2009; Wu et

56   al., 2019). As a result, the microbial water quality of wastewater can be a useful indicator of an

57   area's biological contamination, with outbreaks of several diseases corresponding to increased

58   wastewater titers of pathogenic etiological agents (Hellmér et al., 2014; Manor et al., 1999;

59   Rothman et al., 2021; Wu et al., 2020). The microbial ecology of wastewater is an important

60   topic, with many studies characterizing the microbes present through culturing, PCR- and

61   sequencing-based methods, and generally rely on targeting specific pathogens or metagenomic

62   shotgun DNA sequencing (Hubeny et al., 2022; Jankowski et al., 2022; Kitajima et al., 2018;

63   Martínez-Puchol et al., 2020). While useful, these studies are unable to capture microbial

64   transcription, which provides information about active microbial processes, instead of the

65   genomic potential of wastewater. Moreover, as many important human and crop/livestock

66   pathogens are RNA viruses (Amoah et al., 2020; Bibby and Peccia, 2013; Symonds et al., 2009),

67   we can monitor the presence and spread of *Ribovira* through untargeted metatranscriptomics.

68   Wastewater RNA sequencing can uncover active microbial interactions and metabolic networks,

69   which may inform us of the public and environmental health of the areas served by a given

70   sewage system (Brumfield et al., 2022; Crits-Christoph et al., 2021; Li et al., 2022; Rothman et

71   al., 2021).

72   Wastewater-based epidemiology (WBE) can inform public health about the presence of

73   pathogens in a population without needing to test individuals in healthcare settings (Bivins et al.,

74   2020; Sims and Kasprzyk-Hordern, 2020). Health agencies have used and WBE to detect the

75   presence of human pathogens such as norovirus, polio, SARS coronaviruses, and a variety of

76   bacteria and protists (Hellmér et al., 2014; Manor et al., 1999; Rothman et al., 2021; Wu et al.,

77   2020). For example, WBE has been heavily used to track and monitor the abundance and spread

78   of SARS-CoV-2 during the ongoing COVID-19 pandemic at various population levels (Achak et

79   al., 2021; Karthikeyan et al., 2021; Nemudryi et al., 2020; Peccia et al., 2020; Rothman et al.,

80   2021; Wu et al., 2020). Furthermore, as disease case counts change longitudinally, multiple time

81   points and RNA sequencing are useful to track not only the presence, but the activity of

82   microorganisms which may provide additional information about pathogens over longer time

83   periods (Faust et al., 2015; Joseph et al., 2019; Marcelino et al., 2019; Nemudryi et al., 2020).

84   Lastly, by broadly sequencing RNA, we may be able to discover new targets for microbial water

85   quality assays in order to detect and monitor for sewage contamination of the environment and

86   water sources (Cao et al., 2015; Farkas et al., 2019; Jiang et al., 2022; Kitajima et al., 2018;

87   Zimmer-Faust et al., 2021).

88   Antimicrobial resistance (AMR) is a worldwide concern that inhibits effective treatment

89   of disease and increases healthcare burden and morbidity of infections (World Health

90   Organization, 2021). Wastewater contains a complex diversity of AMR genes, which allows for

91   horizontal gene transfer (HGT) between antimicrobial resistant organisms and those species or

92   strains that are currently susceptible to antimicrobial therapies (Joseph et al., 2019; Ju et al.,

93   2019; Sims and Kasprzyk-Hordern, 2020). As AMR and HGT are important to monitor for

94   public and agricultural health, many studies have employed sequencing and targeted PCR-based

95   technologies to assay the AMR genomic content of wastewater (Karkman et al., 2018). While

96   useful, these studies typically rely on DNA-based technologies which cannot measure the

97   transcriptional activity of these genes or the organisms that harbor AMR, and may better indicate

98   the severity and abundance of antimicrobial resistant infections across a population (de Nies et

99   al., 2021; Ju et al., 2019; Marcelino et al., 2019). By employing RNA-sequencing, we are better

100   able to understand the disease ecology and AMR activity of wastewater-inhabiting organisms

101   and those deposited through the waste stream, and the specific mutations that cause AMR

102   (Alcock et al., 2020). Lastly, through careful sampling, changes in AMR transcription can be

103   tracked over time, likely providing finer-scale information about the severity and seasonality of

104   AMR infections during the ongoing COVID-19 pandemic (Langford et al., 2020; Rose et al.,

105   2021).

106   Studying wastewater microbial ecology and tracking the activity of disease-associated

107   microbes and AMR is vital to public health and environmental monitoring. In this study, we used

108   metatranscriptomic sequencing to characterize the RNA world of 275 samples across eight

109   wastewater treatment plants (WTPs) representing approximately 16 million people across

110   Southern California. We investigated several lines of inquiry: First, what is the transcriptomic

111   diversity of microorganisms in Southern California wastewater, and does it vary longitudinally?

112   Second, what AMR genes are being actively transcribed in wastewater? Third, are there

113   conserved biochemical pathways across wastewater, and does this metabolic potential vary?

114   Lastly, are there largescale patterns of microbial transcription in Southern California's

115   wastewater, and is there a temporal component to any of these patterns?

116 2. Materials and Methods:

117 *2.1 Sample collection*

118   We previously reported the sample collection and handling procedure in Rothman et al

119 2021 (Rothman et al., 2021), and note that the viromes of 94 samples were previously reported in

120 that study. Briefly, we collected 275 1-liter 24-hour composite influent wastewater samples by

121 autosampler at eight WTPs across Southern California between August 2020 – August 2021

122 (Table 1). We aliquoted and stored 50 mL of sample at 4 °C until RNA extraction.

123 *2.2 RNA extraction and sequencing library preparation*

124   We used a protocol based on Crits-Christoph 2021 (Crits-Christoph et al., 2021) and Wu

125 et al 2020 (Wu et al., 2020), in which we pasteurized 50 mL of influent wastewater in a 65 °C

126 water bath for 90 minutes, then filtered samples through a sterile 0.22-µM filter (VWR, Radnor,

127 PA). We then centrifuged the sample at 3,000 xg through 10-kDa Amicon filters

128 (MilliporeSigma, Burlington, MA) and stored the concentrate at -80 °C until RNA extraction.

129 We then used an Invitrogen PureLink RNA Mini Kit with added DNase step (Invitrogen,

130 Waltham, MA) following the manufacturer's instructions to extract RNA and stored the resulting

131 RNA at -80 °C until library preparation.

132   The University of California Irvine Genomics High Throughput Facility (GHTF) handled

133 all library preparation steps. Briefly, the GHTF used the Illumina RNA Prep for Enrichment kit

134 (Illumina, San Diego, CA) on each RNA sample, then sequenced the paired end libraries as 2 x

135 100bp or 2 x 150 bp (supplemental file SF1) with an S4 300 cycle kit on an Illumina NovaSeq

136 6000 over four batches.

137 *2.3 Bioinformatics and data processing*

138   We received the data from the GHTF as demultiplexed FASTQ files and used the UCI

139   High Performance Community Computing Cluster for data processing. We used BBTools

140   "bbduk" (Bushnell, 2014) to remove adapters, low-quality bases, and primers, then removed

141   PCR duplicates with BBTools "dedupe." After deduplication, we removed reads mapping to the

142   human genome (hg38) with Bowtie2 (Langmead and Salzberg, 2012), then used Kraken2 (Wood

143   et al., 2019) and Bracken (Lu et al., 2017) databases built with the NCBI RefSeq database of

144   bacteria, archaea, and viruses (January 2021), to taxonomically classify reads. We then tabulated

145   these reads and used these tables for downstream diversity analysis (supplemental file 2).

146   For community analyses, we normalized the transcript reads into within-sample relative

147   abundances in R, removed reads corresponding to less than 0.01% relative abundance, then

148   calculated Shannon Diversity indices and Bray-Curtis dissimilarity matrices with the R package

149   "vegan" (Oksanen et al., 2017). We generated nonmetric multidimensional scaling (NMDS)

150   ordinations, then tested the diversity metrics for significant differences with Kruskal-Wallis tests

151   (alpha diversity) and Adonis PERMANOVA (beta diversity) with "vegan." We assessed the

152   relationship of diversity with time with linear mixed effects models (lmer) in the R package

153   "lmerTest" using WTP and sequencing batch as random effects (Kuznetsova et al., 2017), and

154   plotted all diversity analyses with "ggplot2" (Wickham, 2009), "ggrepel" (Slowikowski, 2018),

155   and "patchwork" (Pedersen, 2020). Because we collected samples from Escondido, Hyperion,

156   and Point Loma for a much longer period of time than the other WTPs, we ran the above

157   analyses two ways: all WTPs together from August – November 2020, and Escondido, Hyperion,

158   and Point Loma samples for the full year separately.

159   We used HUMAnN3 (Beghini et al., 2021) with default settings to assign functional gene

160   pathway annotations to reads using the UniRef90 (Suzek et al., 2015) and Metacyc (Caspi et al.,

161   2020) databases. We also used RGI (the Resistance Gene Identifier) and the CARD and

162   WildCARD databases (Alcock et al., 2020) to assign predicted antimicrobial resistance ontology

163   identities (AROs) to the reads, then normalized all pathway abundances and AMR gene identities

164   to transcripts per million (TPM). We compared microbial abundances, pathway abundances, and

165   AMR gene abundances at greater than 0.01% relative abundance and present in 50% of samples

166   between WTPs with ANCOM2.1 using sample collection month as an adjustment for covariates

167   and sequencing batch as a random effect in the ANCOM models. We then plotted $\log_{10}$

168   transformed counts of significantly differentially abundant viruses, bacterial genera, and AMR

169   genes on a heatmap allowing the taxa to cluster with the Ward D2 algorithm with the R package

170   "pheatmap" (Kolde, 2019). We used MaAsLin2 (Mallick et al., 2021) for longitudinal analyses

171   of the above-mentioned variables, and included WTP and sequencing batch as random effects in

172   the models, and we adjusted ANCOM and MaAsLin2 statistical tests for multiple comparisons

173   with the Benjamini-Hochberg correction. We report the linear model coefficient with time of

174   MaAsLin2 analyses on each plot and refer to Zenodo (doi.org/10.5281/zenodo.6829029)

175   (Rothman et al., 2022) for individual scatterplots.

176   *2.4 Data and code availability*

177   Representative analyses scripts and code are available at

178   github.com/jasonarothman/wastewater_metatranscriptomics_socal_aug20_aug21 and raw

179   sequencing files have been deposited at the NCBI Sequence Read Archive under accession

180   numbers PRJNA729801. Data tables containing taxa abundances, HUMAnN3 pathway

181   annotations, and RGI assigned predicted antimicrobial resistance ontology identities are

182   available as a Dryad dataset (https://doi.org/10.7280/D11Q30) (Rothman et al., 2022)

183   3. Results:

184   *3.1 Library statistics and microbial sample composition*

185   We obtained a total of 4,336,566,730 quality-filtered, nonhuman, paired-end reads across

186   275 samples from eight WTPs (average: 15,769,334 reads per sample, range: 1,039,430 –

187   88,651,858). With Kraken2, we classified an average of 55.0% of our reads (range 8.7 – 83.5%),

188   of which an average of 48.1% (range 7.0 – 83.0%) were bacterial, 0.2% were archaeal (range

189   0.02 – 3.8%), and 5.9% (range 0.03 – 38.3%) were viral (Fig. 2). Due to the low relative

190   abundance of archaea and known questionable classification accuracy, we chose to focus on

191   bacteria and viruses for diversity analyses.

192   We detected transcripts from a total of 6,449 bacterial and 6,888 viral species across all

193   samples, however due to the likelihood of the taxonomic classifier reporting spurious species, we

194   removed species accounting for < 0.01% average relative abundance within each domain. This

195   filtering left us with 935 bacterial and 134 viral species present, which we used for downstream

196   analyses. We also tabulated 245 bacterial families present in the same fashion as above. Because

197   we had an uneven longitudinal distribution of samples, we analyzed diversity, differential

198   abundance, and longitudinal relationships in two ways: First, samples where we had all eight

199   WTPs were analyzed together representing N = 98, covering the months of August – November

200   2020. Second, we analyzed samples from Escondido, Hyperion, and Point Loma WTPs, where

201   we had an entire year of sampling (N = 214), covering August 2020 – August 2021.

202   *3.2 Antimicrobial resistance transcription*

203   We detected transcripts matching 2,128 unique antibiotic resistance ontology identifiers

204   (AROs) through use of RGI and the CARD database (Fig. 2, Dryad:

205   https://doi.org/10.7280/D11Q30). AMR alpha diversity between August – November 2020

206   significantly differed between WTPs ($H_{(7)} = 33.7$, $P < 0.001$), but not over time (t = -0.3, P =

207   0.74). AMR beta diversity during this time only differed between WTPs ($P < 0.001$, $R^2 = 0.43$),

208   and not by month ($P = 0.08$, $R^2 = 0.03$), an interaction of WTP and month ($P = 0.10$, $R^2 = 0.16$),

209   or sequencing batch ($P = 0.05$, $R^2 = 0.02$), and slightly changed over time (t = -2.3, P = 0.03)

210   (Fig. 3). Several AMR transcripts were differentially abundant between WTPs, and for easier

211   discrimination between the categories, we separated them into ribosomal RNA mutations and

212   non-rRNA AMR genes: 29 rRNA AMR mutants (W > 88, $P_{adj} < 0.05$) and 17 non-rRNA genes

213   (W > 140, $P_{adj} < 0.05$) differed between WTPs (Fig. 3, supplemental file SF2).

214   When considering the entire year, AMR alpha diversity differed between WTPs ($H_{(2)} =$

215   28.6, $P < 0.001$), but not over time (t = -0.6, P = 0.68). AMR beta diversity differed between

216   WTPs ($P < 0.001$, $R^2 = 0.13$), month ($P < 0.001$, $R^2 = 0.16$), an interaction between WTP and

217   month ($P = 0.007$, $R^2 = 0.12$), with significant batch effects ($P < 0.001$, $R^2 = 0.06$), and again,

218   changed over time (t = 3.3, P = 0.001) (Fig. 3). We considered AMR transcripts from rRNA

219   genes and non-ribosomal genes separately as above. For rRNA genes, we found that 26

220   positively and 13 negatively correlated with time ($P_{adj} < 0.05$), while 45 did not, and for non-

221   ribosomal genes, 38 positively and 1 negatively changed over time ($P_{adj} < 0.05$), while 256 did

222   not change significantly (Fig. 3, supplemental file SF3, Zenodo:

223   doi.org/10.5281/zenodo.6829029).

224   *3.3 Bacterial transcriptional ecology*

225   We found that the top ten most proportionally abundant bacterial families represented an

226   average of 58.6% (range 17.7 – 82.3%) of bacterial transcripts. These families (in descending

227   average proportional abundance) were: Campylobacteraceae, Pseudomonadaceae,

228   Enterobacteriaceae, Neisseriaceae, Moraxellaceae, Comamonadaceae, Burkholderiaceae,

229   Aeromonadaceae, Weeksellaceae, and Methylobacteriaceae (Fig. 2).

230   From August – November 2020, bacterial transcript alpha diversity significantly differed

231   between WTP ($H_{(7)}$ = 55.5, $P < 0.001$), but not over time (t = -1.3, P = 0.22). Bacterial beta

232   diversity was significantly different across WTPs ($P < 0.001$, R^2 = 0.30) and month ($P < 0.001$,

233   R^2 = 0.09), with no interaction between WTP and month (P = 0.13, R^2 = 0.16), was affected by

234   sequencing batch ($P < 0.001$, R^2 = 0.03), and changed over time (t = -2.4, P = 0.02). We also

235   found that transcripts from 222/564 bacterial genera were significantly differentially abundant

236   between WTPs during this time period (W > 507, $P_{adj} < 0.05$, Fig. 4, supplemental file SF2).

237   Across the entire year, alpha diversity was not different between WTPs ($H_{(2)}$ = 1.1, P =

238   0.59), and did not differ over time (t = 1.6, P = 0.12). Beta diversity differed between WTP (P <

239   0.001, R^2 = 0.07), month ($P < 0.001$, R^2 = 0.20), and the interaction of WTP and month (P =

240   0.002, R^2 = 0.11) with significant batch effects ($P < 0.001$, R^2 = 0.06), and over time as a

241   continuous variable (t = 4.8, $P < 0.001$). We tracked the transcription of bacterial genera across

242   the year, and found that 172 genera increased, 63 genera decreased, and 295 did not change

243   significantly over time (Fig. 4, supplemental file SF3, Zenodo:

244   doi.org/10.5281/zenodo.6829029).

245   *3.4 Viral ecology*

246   We did not group viruses by family because of the dominance of Virgaviridae, and

247   instead report summary statistics of the ten most proportionally abundant viral species as this

248   provides more information. These viruses represented an average proportional viral abundance of

249   92.4% (range 33.1 - 99.5%; in descending average proportional abundance): Tomato brown

250   rugose fruit virus, Cucumber green mottle mosaic virus, Pepper mild mottle virus, crAssphage,

251   Tomato mosaic virus, Tropical soda apple mosaic virus, Tobacco mild green mosaic virus,

252   Tomato mottle mosaic virus, Melon necrotic spot virus, and Pseudomonas virus PMBT3 (Fig. 2).

253   Over August – November 2020, viral alpha diversity differed between WTPs ($H_{(7)}$ = 35.1,

254   P < 0.001), but not over time (t = -0.57, P = 0.58). Beta diversity differed between WTPs (P <

255   0.001, R^2 = 0.31), by month (P = 0.003, R^2 = 0.07), but not by an interaction between WTP and

256   month (P = 0.69, R^2 = 0.13), by batch (P = 0.07, R^2 = 0.02), or over time (t = -1.6, P = 0.11).

257   During this time period, only 11 viruses were differentially abundant between WTPs (Fig. 5,

258   supplemental file SF2).

259   The full-year samples showed significantly different alpha diversity between WTP ($H_{(2)}$ =

260   55.4, P < 0.001) but not over time (t = 0.11, P = 0.91). Long-term beta diversity differed between

261   WTPs (P = 0.005, R^2 = 0.03), month (P = 0.001, R^2 = 0.11), with no interaction between WTP

262   and month (P = 0.34, R^2 = 0.09), with significant batch effects (P = 0.001, R2 = 0.11), and

263   changed significantly over time (t = 4.3, P < 0.001). When considering the proportional

264   abundance of individual virus species over the year, 22 viruses increased 16 decreased, and 102

265   did not change over time (Fig. 5, supplemental file SF3, doi.org/10.5281/zenodo.6829029).

266   *3.5 Metabolic pathway transcription*

267   Across samples that successfully processed through HUMAnN3 (N = 252), we detected

268   transcripts that mapped to 474 Metacyc metabolic pathways (Dryad:

269   https://doi.org/10.7280/D11Q30). Most commonly, we found transcriptional activity from

270   pathways such as nucleotide biosynthesis, ubiquitination, amino acid biosynthesis, and central

271   carbon metabolism, while we also detected rarer pathways involved in the degradation of

272   xenobiotics including toluene, atrazine, nitrobenzoate, and octane.

273   Metabolic transcript alpha diversity was not significantly different across WTPs from

274   August – November 2020 ($H_{(7)}$ = 4.8, P = 0.68) and did not change over time (t = 1.2, P = 0.222).

275   Likewise, metabolic transcript beta diversity during this period was not different between WTPs

276   (P = 0.18, R^2 = 0.09), but slightly differed between months (P = 0.003, R^2 = 0.07) with an

277   interaction between month and WTP (P = 0.03, R^2 = 0.24) (Fig. 6), with significant batch effects

278   (P < 0.001, R^2 = 0.06), but did not change over time (t = 0.9, P = 0.38). There were no

279   differentially-expressed metabolic pathways across WTPs during this time period.

280   Across the full year, transcript alpha diversity differed between WTPs ($H_{(2)}$ = 14.4, P <

281   0.001), but not over time (t = 1.3, P = 0.19). Beta diversity slightly differed between WTPs (P =

282   0.008, R^2 = 0.02), month (P < 0.001, R^2 = 0.17), with an interaction between WTP and month (P

283   = 0.005, R^2 = 0.12) with significant sequencing batch effects (P = 0.002, R^2 = 0.04), and did not

284   change longitudinally (t = 0.8, P = 0.41). The transcription of few metabolic pathways had a

285   significant association with time, as only 12 were positively, and one was negatively correlated,

286   out of 205 pathways total (Fig. 6, supplemental file SF3, Zenodo:

287   doi.org/10.5281/zenodo.6829029).

288   4. Discussion:

289   Composite wastewater samples from Southern California over the year contained RNA

290   transcripts derived from a wide diversity of microorganisms. To the best of our knowledge, our

291   study representing a sewer shed of 16 million people is the most complete metatranscriptomic

292   characterization of a large metropolitan region's wastewater to date. Most notably, we show

293   evidence of actively transcribed antimicrobial resistance (AMR) genes that encode resistance to a

294   variety of commonly-administered antimicrobial drugs including macrolides, aminoglycosides,

295   tetracycline and other AMR classes (Alcock et al., 2020). Likewise, we also show that bacterial

296   transcription and RNA viral diversity differed between wastewater treatment plants (WTPs), and

297   that sequencing wastewater RNA can be a useful tool for wastewater-based epidemiology

298   (WBE) (Brumfield et al., 2022; Crits-Christoph et al., 2021; de Nies et al., 2021; Rothman et al.,

299   2021; Xagoraraki and O'brien, 2020). Finally, we examined the total RNA pool and described

300   metabolic pathway transcription to show that wastewater metabolism is largely consistent across

301   WTPs and over time, but that there are slight signatures of geographical location (Gulino et al.,

302   2020). Our results suggest that RNA sequencing is a viable tool to understand the complex

303   matrix that wastewater represents and is useful in assaying the microbes associated with large

304   populations.

305   *4.1 Antimicrobial resistance transcription across Southern California wastewater*

306   Wastewater is known to harbor an array of AMR genes, and several studies have

307   sequenced and/or quantified many of these genes in wastewater (de Nies et al., 2021; Ju et al.,

308   2019; Raza et al., 2022; Yin et al., 2021). Our study differs in that we demonstrate transcriptional

309   activity through RNA-sequencing, rather than the genomic potential of the sampled organisms.

310   We found a wide diversity of transcribed AMR genes in our data, including components of the

311   multidrug efflux pumps adeFGH (Coyne et al., 2010) and its repressor acrS (Hirakawa et al.,

312   2008), the gene tetQ (Nikolich et al., 1992), which encodes a ribosomal protection protein

313   against tetracycline, *Staphylococcus aureus's* multidrug efflux protein lmrS (Floyd et al., 2010),

314   genes in the aminoglycoside resistance series aadA and aph(3") (Ramirez and Tolmasky, 2010),

315   and several variants of the glycopeptide resistance gene vanR (Courvalin, 2006). Many of these

316   transcripts have been previously detected in WTPs, or in animals that resided in wastewater

317   (Brumfield et al., 2022; Marcelino et al., 2019). Because we did not deplete rRNAs during

318   library preparations, most of our bacterial transcripts were ribosomal RNAs. We detected rRNA

319   mutations that confer macrolide resistance in the medically important taxa *Neisseria,*

320   *Campylobacter, Salmonella, Helicobacter, Staphylococcus, Streptococcus, Klebsiella*, and many

321   others. These genera (and subsequent AMR-resistant rRNAs) were ubiquitous in our samples

322   and are often found in wastewater (Jankowski et al., 2022; Joseph et al., 2019; Ju et al., 2019).

323   Our results indicate transcriptional evidence of widespread AMR activity, and we posit that this

324   AMR presence is likely to be found in other wastewater catchments making metatranscriptomics

325   useful for tracking AMR across wide areas (de Nies et al., 2021). The diversity of AMR genes in

326   our samples differed between WTPs, and there were a few AMR genes differentially abundant

327   between WTPs – mostly mutant rRNAs. This finding supports studies that show geographic

328   differences between AMR (Raza et al., 2022; Yin et al., 2021), but there are likely other factors

329   impacting the diversity of AMR, such as disease load in the served populations. Interestingly, we

330   noticed a general increase over time in the proportional abundance of several transcripts from the

331   major facilitator superfamily (MFS) and resistance-nodulation-cell division (RND) antibiotic

332   efflux pumps - which are often implicated in multidrug resistance (Li and Nikaido, 2009) - along

333   with beta-lactamases, and aminoglycoside/macrolide resistant rRNAs (Alcock et al., 2020).

334   These data support studies showing an increase in antibiotic resistance (Ju et al., 2019) and the

335   prevalence of AMR genes, but may also be impacted by seasonal changes in the waste stream

336   (Yang et al., 2013). Likewise, as antibiotic use has risen during the COVID-19 pandemic

337   (Langford et al., 2020; Rose et al., 2021), we may be observing a concurrent rise in AMR

338   transcription in wastewater, although because our samples were solely collected during COVID-

339   19, we are only able to speculate.

340   *4.2 Viral ecology of Southern California wastewater*

341   Plant-infecting tobamoviruses dominated the viromes of our samples regardless of source

342   or time of year (Bačnik et al., 2020; Brumfield et al., 2022; Cantalupo et al., 2011; Crits-

343   Christoph et al., 2021), although we also found substantial numbers of reads mapping to phages

344   including crAssphage and assorted bacteriophages. While most known phages have DNA

345   genomes, previous studies have identified phages in wastewater RNA (Crits-Christoph et al.,

346   2021; Wilder et al., 2021). We may be detecting novel RNA viruses, or transcription of either

347   DNA or RNA based phage genomes. Viral diversity differed when tested across all WTPs and

348   over the full year, supporting studies that suggest geographical signatures of viruses in

349   wastewater, and may be due to differences in human diet and viral excretion, along with disease

350   dynamics in bacteria and/or eukaryotic hosts (Bibby and Peccia, 2013; Brumfield et al., 2022;

351   Gulino et al., 2020). Likewise, several viruses were differentially abundant over time, which may

352   be due to underlying infection trends or due to unknown seasonality effects (Brinkman et al.,

353   2017; Kazama et al., 2016). While overall viral diversity was different between WTPs and

354   changed over time, highly abundant viruses tended to be present in most samples, which may

355   afford new targets in establishing microbial water quality or the detection of sewage pollution

356   (Cao et al., 2015; Jiang et al., 2022; Kitajima et al., 2018). Similarly, we detected several human-

357   infecting viruses (i.e. Norwalk Virus and SARS-CoV-2) which provides support for WBE efforts

358   (Crits-Christoph et al., 2021; Nemudryi et al., 2020; Rothman et al., 2021, 2020; Xagoraraki and

359   O'brien, 2020), and we suggest that RNA sequencing of wastewater should be used in

360   conjunction with targeted and quantificational approaches to assist in passively monitoring

361   diseases across large populations.

362   *4.3 Bacterial ecology and metabolic pathways in Southern California wastewater*

363   Similar to other studies, we detected transcripts from bacterial species in wastewater -

364   mostly in the form of rRNA reads (de Nies et al., 2021; Joseph et al., 2019). Human pathogens

365   were broadly represented in our data, including ESKAPE bacteria (*Enterococcus faecium,*

366   *Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas*

367   *aeruginosa,* and *Enterobacter* spp.), *Campylobacter jejuni*, *Salmonella* spp., *Helicobacter pylori*,

368   *Haemophilus* spp., sexually transmitted infectious (STIs) agents, and bacteria commonly found

369   in the environment. Much as with viruses, the bacterial profiles of WTPs were different,

370   although many species were ubiquitous throughout the samples (Wu et al., 2019). There were

371   also noticeable changes in the relative proportional transcript abundance over time, with many

372   bacterial genera displaying a bimodal periodicity: Higher transcript abundance during Winter

373   and Summer, and generally higher as time proceeded from August 2020 to August 2021. Other

374   work has shown a distinct seasonality to the wastewater microbial community (Peces et al.,

375   2022) - and our data supports this as well - although there are many other factors that can affect

376   wastewater communities, such as pH, flux, dissolved oxygen, and detergents (Wu et al., 2019).

377   Likewise, we recognize that our RNA extraction methods were harsh, and surely resulted in

378   nucleic acid degradation, which likely affects the accuracy of our results (Schuierer et al., 2017).

379   Non-ribosomal bacterial metabolism was apparent in our data with transcripts mapping to

380   widely-conserved pathways such as nucleotide and amino acid biosynthesis and ubiquitination,

381   with no pathways differing between WTPs or over time (Caspi et al., 2020). Collectively, our

382   results suggest that sequencing bacterial species and their constituent metabolic pathways

383   common to wastewater may be useful for monitoring disease through WBE, and that novel

384   targets to assay microbial water quality may be possible.

385   5. Conclusion:

386   In our opinion, this large-scale longitudinal dataset represents an unprecedented

387   metatranscriptomic characterization of wastewater across a large population and region. We

388   detected a wide diversity of transcribed AMR genes, suggesting that RNA sequencing is a

389   powerful tool for WBE and may be useful in monitoring the spread and intensity of AMR.

390   Within our study, we sequenced the viromes of a large portion of Southern California's

391   wastewater catchment area and show that plant-infecting viruses dominate the RNA viral

392   fraction, which may have additional uses in detecting agricultural disease outbreaks. Similarly,

393   we detected numerous human pathogens and observed changes in the relative proportions of

394   these taxa, lending more credence to WBE as a vital component to public health and microbial

395   water quality assays. We suggest that future transcriptomic studies target disease-causing taxa in

396   wastewater to understand and refine WBE and its usefulness to human health more deeply.

397

408   and 1S10OD021718-01, and access to computing resources from the UCI High Performance

409   Cloud Computing Center.

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 resistome 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/es305181x

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 Gurol, 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 collaborative to maximize contributions in the fight against COVID-19. Environmental Science and Technology. 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.

456   Microbiome analysis for wastewater surveillance during COVID-19. MBio e0059122.
457   https://doi.org/10.1128/mbio.00591-22
458 Bushnell, B., 2014. BBTools software package.
459 Cantalupo, P.G., Calgua, B., Zhao, G., Hundesa, A., Wier, A.D., Katz, J.P., Grabe, M., Hendrix,
460   R.W., Girones, R., Wang, D., Pipas, J.M., 2011. Raw sewage harbors diverse viral
461   populations. MBio 2. https://doi.org/10.1128/mBio.00180-11
462 Cao, Y., Raith, M.R., Griffith, J.F., 2015. Droplet digital PCR for simultaneous quantification of
463   general and human-associated fecal indicators for water quality assessment. Water Res.
464   70, 337–349. https://doi.org/10.1016/j.watres.2014.12.008
465 Caspi, R., Billington, R., Keseler, I.M., Kothari, A., Krummenacker, M., Midford, P.E., Ong,
466   W.K., Paley, S., Subhraveti, P., Karp, P.D., 2020. The MetaCyc database of metabolic
467   pathways and enzymes - a 2019 update. Nucleic Acids Res. 48, D445–D453.
468   https://doi.org/10.1093/nar/gkz862
469 Courvalin, P., 2006. Vancomycin resistance in gram-positive cocci. Clin. Infect. Dis. 42 Suppl 1,
470   S25-34. https://doi.org/10.1086/491711
471 Coyne, S., Rosenfeld, N., Lambert, T., Courvalin, P., Périchon, B., 2010. Overexpression of
472   resistance-nodulation-cell division pump AdeFGH confers multidrug resistance in
473   *Acinetobacter baumannii*. Antimicrob. Agents Chemother. 54, 4389–4393.
474   https://doi.org/10.1128/AAC.00155-10
475 Crits-Christoph, A., Kantor, R.S., Olm, M.R., Whitney, O.N., Al-Shayeb, B., Lou, Y.C.,
476   Flamholz, A., Kennedy, L.C., Greenwald, H., Hinkle, A., Hetzel, J., Spitzer, S., Koble, J.,
477   Tan, A., Hyde, F., Schroth, G., Kuersten, S., Banfield, J.F., Nelson, K.L., 2021. Genome
478   Sequencing of sewage detects regionally prevalent SARS-CoV-2 variants. MBio 12.
479   https://doi.org/10.1128/mBio.02703-20
480 de Nies, L., Busi, S.B., Kunath, B.J., May, P., Wilmes, P., 2021. Mobilome-driven segregation of
481   the resistome in biological wastewater treatment. bioRxiv.
482   https://doi.org/10.1101/2021.11.15.468621
483 Edwards, R.A., Vega, A.A., Norman, H.M., Ohaeri, M., Levi, K., Dinsdale, E.A., Cinek, O.,
484   Aziz, R.K., McNair, K., Barr, J.J., Bibby, K., Brouns, S.J.J., Cazares, A., de Jonge, P.A.,
485   Desnues, C., Díaz Muñoz, S.L., Fineran, P.C., Kurilshikov, A., Lavigne, R., Mazankova,
486   K., McCarthy, D.T., Nobrega, F.L., Reyes Muñoz, A., Tapia, G., Trefault, N., Tyakht,
487   A.V., Vinuesa, P., Wagemans, J., Zhernakova, A., Aarestrup, F.M., Ahmadov, G.,
488   Alassaf, A., Anton, J., Asangba, A., Billings, E.K., Cantu, V.A., Carlton, J.M., Cazares,
489   D., Cho, G.-S., Condeff, T., Cortés, P., Cranfield, M., Cuevas, D.A., De la Iglesia, R.,
490   Decewicz, P., Doane, M.P., Dominy, N.J., Dziewit, L., Elwasila, B.M., Eren, A.M.,
491   Franz, C., Fu, J., Garcia-Aljaro, C., Ghedin, E., Gulino, K.M., Haggerty, J.M., Head,
492   S.R., Hendriksen, R.S., Hill, C., Hyöty, H., Ilina, E.N., Irwin, M.T., Jeffries, T.C., Jofre,
493   J., Junge, R.E., Kelley, S.T., Khan Mirzaei, M., Kowalewski, M., Kumaresan, D., Leigh,
494   S.R., Lipson, D., Lisitsyna, E.S., Llagostera, M., Maritz, J.M., Marr, L.C., McCann, A.,
495   Molshanski-Mor, S., Monteiro, S., Moreira-Grez, B., Morris, M., Mugisha, L., Muniesa,
496   M., Neve, H., Nguyen, N.-P., Nigro, O.D., Nilsson, A.S., O'Connell, T., Odeh, R.,
497   Oliver, A., Piuri, M., Prussin, A.J., Ii, Qimron, U., Quan, Z.-X., Rainetova, P., Ramírez-
498   Rojas, A., Raya, R., Reasor, K., Rice, G.A.O., Rossi, A., Santos, R., Shimashita, J.,
499   Stachler, E.N., Stene, L.C., Strain, R., Stumpf, R., Torres, P.J., Twaddle, A., Ugochi
500   Ibekwe, M., Villagra, N., Wandro, S., White, B., Whiteley, A., Whiteson, K.L.,
501   Wijmenga, C., Zambrano, M.M., Zschach, H., Dutilh, B.E., 2019. Global

502   phylogeography and ancient evolution of the widespread human gut virus crAssphage.
503   Nat Microbiol 4, 1727–1736. https://doi.org/10.1038/s41564-019-0494-6
504 Farkas, K., Adriaenssens, E.M., Walker, D.I., McDonald, J.E., Malham, S.K., Jones, D.L., 2019.
505   Critical evaluation of crassphage as a molecular marker for human-derived wastewater
506   contamination in the aquatic environment. Food Environ. Virol. 11, 113–119.
507   https://doi.org/10.1007/s12560-019-09369-1
508 Faust, K., Lahti, L., Gonze, D., de Vos, W.M., Raes, J., 2015. Metagenomics meets time series
509   analysis: unraveling microbial community dynamics. Curr. Opin. Microbiol. 25, 56–66.
510   https://doi.org/10.1016/j.mib.2015.04.004
511 Floyd, J.L., Smith, K.P., Kumar, S.H., Floyd, J.T., Varela, M.F., 2010. LmrS is a multidrug
512   efflux pump of the major facilitator superfamily from *Staphylococcus aureus*.
513   Antimicrob. Agents Chemother. 54, 5406–5412. https://doi.org/10.1128/AAC.00580-10
514 Gulino, K., Rahman, J., Badri, M., Morton, J., Bonneau, R., Ghedin, E., 2020. Initial mapping of
515   the New York City wastewater virome. mSystems 5.
516   https://doi.org/10.1128/mSystems.00876-19
517 Hellmér, M., Paxéus, N., Magnius, L., Enache, L., Arnholm, B., Johansson, A., Bergström, T.,
518   Norder, H., 2014. Detection of pathogenic viruses in sewage provided early warnings of
519   hepatitis A virus and norovirus outbreaks. Appl. Environ. Microbiol. 80, 6771–6781.
520   https://doi.org/10.1128/AEM.01981-14
521 Hirakawa, H., Takumi-Kobayashi, A., Theisen, U., Hirata, T., Nishino, K., Yamaguchi, A.,
522   2008. AcrS/EnvR represses expression of the acrAB multidrug efflux genes in
523   *Escherichia coli*. J. Bacteriol. 190, 6276–6279. https://doi.org/10.1128/JB.00190-08
524 Hubeny, J., Korzeniewska, E., Buta-Hubeny, M., Zieliński, W., Rolbiecki, D., Harnisz, M.,
525   2022. Characterization of carbapenem resistance in environmental samples and
526   *Acinetobacter* spp. isolates from wastewater and river water in Poland. Sci. Total
527   Environ. 822, 153437. https://doi.org/10.1016/j.scitotenv.2022.153437
528 Jankowski, P., Gan, J., Le, T., McKennitt, M., Garcia, A., Yanaç, K., Yuan, Q., Uyaguari-Diaz,
529   M., 2022. Metagenomic community composition and resistome analysis in a full-scale
530   cold climate wastewater treatment plant. Environ Microbiome 17, 3.
531   https://doi.org/10.1186/s40793-022-00398-1
532 Jiang, S.C., Bischel, H.N., Goel, R., Rosso, D., Sherchan, S.P., Whiteson, K.L., Yan, T., Solo-
533   Gabriele, H.M., 2022. Integrating virus monitoring strategies for safe non-potable water
534   reuse. Water 14, 1187. https://doi.org/10.3390/w14081187
535 Joseph, S.M., Battaglia, T., Maritz, J.M., Carlton, J.M., Blaser, M.J., 2019. Longitudinal
536   comparison of bacterial diversity and antibiotic resistance genes in New York City
537   sewage. mSystems 4. https://doi.org/10.1128/mSystems.00327-19
538 Ju, F., Beck, K., Yin, X., Maccagnan, A., McArdell, C.S., Singer, H.P., Johnson, D.R., Zhang,
539   T., Bürgmann, H., 2019. Wastewater treatment plant resistomes are shaped by bacterial
540   composition, genetic exchange, and upregulated expression in the effluent microbiomes.
541   ISME J. 13, 346–360. https://doi.org/10.1038/s41396-018-0277-8
542 Karkman, A., Do, T.T., Walsh, F., Virta, M.P.J., 2018. Antibiotic-resistance genes in waste
543   water. Trends Microbiol. 26, 220–228. https://doi.org/10.1016/j.tim.2017.09.005
544 Karthikeyan, S., Ronquillo, N., Belda-Ferre, P., Alvarado, D., Javidi, T., Longhurst, C.A.,
545   Knight, R., 2021. High-throughput wastewater SARS-CoV-2 detection enables
546   forecasting of community infection dynamics in San Diego County. mSystems 6.
547   https://doi.org/10.1128/mSystems.00045-21

548 Kazama, S., Masago, Y., Tohma, K., Souma, N., Imagawa, T., Suzuki, A., Liu, X., Saito, M.,
549   Oshitani, H., Omura, T., 2016. Temporal dynamics of norovirus determined through
550   monitoring of municipal wastewater by pyrosequencing and virological surveillance of
551   gastroenteritis cases. Water Res. 92, 244–253.
552   https://doi.org/10.1016/j.watres.2015.10.024
553 Kitajima, M., Sassi, H.P., Torrey, J.R., 2018. Pepper mild mottle virus as a water quality
554   indicator. npj Clean Water 1, 19. https://doi.org/10.1038/s41545-018-0019-5
555 Kolde, R., 2019. pheatmap: Pretty Heatmaps.
556 Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. {lmerTest} Package: Tests in
557   Linear Mixed Effects Models. J. Stat. Softw. 82, 1–26.
558   https://doi.org/10.18637/jss.v082.i13
559 Langford, B.J., So, M., Raybardhan, S., Leung, V., Westwood, D., MacFadden, D.R., Soucy, J.-
560   P.R., Daneman, N., 2020. Bacterial co-infection and secondary infection in patients with
561   COVID-19: a living rapid review and meta-analysis. Clin. Microbiol. Infect. 26, 1622–
562   1629. https://doi.org/10.1016/j.cmi.2020.07.016
563 Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9,
564   357–359. https://doi.org/10.1038/nmeth.1923
565 Li, R., Zhu, L., Cui, L., Zhu, Y.-G., 2022. Viral diversity and potential environmental risk in
566   microplastic at watershed scale: Evidence from metagenomic analysis of plastisphere.
567   Environ. Int. 161, 107146. https://doi.org/10.1016/j.envint.2022.107146
568 Li, X.-Z., Nikaido, H., 2009. Efflux-mediated drug resistance in bacteria: an update. Drugs 69,
569   1555–1623. https://doi.org/10.2165/11317030-000000000-00000
570 Lu, J., Breitwieser, F.P., Thielen, P., Salzberg, S.L., 2017. Bracken: Estimating species
571   abundance in metagenomics data. PeerJ Computer Science 2017, e104.
572   https://doi.org/10.7717/peerj-cs.104
573 Lu, L., Guest, J.S., Peters, C.A., Zhu, X., Rau, G.H., Ren, Z.J., 2018. Wastewater treatment for
574   carbon capture and utilization. Nature Sustainability 1, 750–758.
575   https://doi.org/10.1038/s41893-018-0187-9
576 Mallick, H., Rahnavard, A., McIver, L.J., Ma, S., Zhang, Y., Nguyen, L.H., Tickle, T.L.,
577   Weingart, G., Ren, B., Schwager, E.H., Chatterjee, S., Thompson, K.N., Wilkinson, J.E.,
578   Subramanian, A., Lu, Y., Waldron, L., Paulson, J.N., Franzosa, E.A., Bravo, H.C.,
579   Huttenhower, C., 2021. Multivariable association discovery in population-scale meta-
580   omics studies. PLoS Comput. Biol. 17, e1009442.
581   https://doi.org/10.1371/journal.pcbi.1009442
582 Manor, Y., Handsher, R., Halmut, T., Neuman, M., Bobrov, A., Rudich, H., Vonsover, A.,
583   Shulman, L., Kew, O., Mendelson, E., 1999. Detection of poliovirus circulation by
584   environmental surveillance in the absence of clinical cases in Israel and the Palestinian
585   Authority. J. Clin. Microbiol. 37, 1670–1675. https://doi.org/10.1128/jcm.37.6.1670-
586   1675.1999
587 Marcelino, V.R., Wille, M., Hurt, A.C., González-Acuña, D., Klaassen, M., Schlub, T.E., Eden,
588   J.-S., Shi, M., Iredell, J.R., Sorrell, T.C., Holmes, E.C., 2019. Meta-transcriptomics
589   reveals a diverse antibiotic resistance gene pool in avian microbiomes. BMC Biol. 17, 31.
590   https://doi.org/10.1186/s12915-019-0649-1
591 Martínez-Puchol, S., Rusiñol, M., Fernández-Cassi, X., Timoneda, N., Itarte, M., Andrés, C.,
592   Antón, A., Abril, J.F., Girones, R., Bofill-Mas, S., 2020. Characterisation of the sewage

593   virome: comparison of NGS tools and occurrence of significant pathogens. Sci. Total
594   Environ. 713, 136604. https://doi.org/10.1016/j.scitotenv.2020.136604
595 McLellan, S.L., Huse, S.M., Mueller-Spitz, S.R., Andreishcheva, E.N., Sogin, M.L., 2010.
596   Diversity and population structure of sewage-derived microorganisms in wastewater
597   treatment plant influent. Environ. Microbiol. 12, 378–392. https://doi.org/10.1111/j.1462-
598   2920.2009.02075.x
599 Nemudryi, A., Nemudraia, A., Wiegand, T., Surya, K., Buyukyoruk, M., Cicha, C.,
600   Vanderwood, K.K., Wilkinson, R., Wiedenheft, B., 2020. Temporal detection and
601   phylogenetic assessment of SARS-CoV-2 in municipal wastewater. Cell Rep Med 1,
602   100098. https://doi.org/10.1016/j.xcrm.2020.100098
603 Newton, R.J., McClary, J.S., 2019. The flux and impact of wastewater infrastructure
604   microorganisms on human and ecosystem health. Curr. Opin. Biotechnol. 57, 145–150.
605   https://doi.org/10.1016/j.copbio.2019.03.015
606 Nikolich, M.P., Shoemaker, N.B., Salyers, A.A., 1992. A *Bacteroides* tetracycline resistance
607   gene represents a new class of ribosome protection tetracycline resistance. Antimicrob.
608   Agents Chemother. 36, 1005–1012. https://doi.org/10.1128/AAC.36.5.1005
609 Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R.,
610   O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H.,
611   2017. vegan: Community Ecology Package.
612 Peccia, J., Zulli, A., Brackney, D.E., Grubaugh, N.D., Kaplan, E.H., Casanovas-Massana, A.,
613   Ko, A.I., Malik, A.A., Wang, D., Wang, M., Warren, J.L., Weinberger, D.M., Arnold,
614   W., Omer, S.B., 2020. Measurement of SARS-CoV-2 RNA in wastewater tracks
615   community infection dynamics. Nat. Biotechnol. 38, 1164–1167.
616   https://doi.org/10.1038/s41587-020-0684-z
617 Peces, M., Dottorini, G., Nierychlo, M., Andersen, K.S., Dueholm, M.K.D., Nielsen, P.H., 2022.
618   Microbial communities across activated sludge plants show recurring species-level
619   seasonal patterns. ISME Communications 2, 1–11. https://doi.org/10.1038/s43705-022-
620   00098-4
621 Pedersen, T.L., 2020. Patchwork: The composer of plots. R package version 1, 182.
622 Ramirez, M.S., Tolmasky, M.E., 2010. Aminoglycoside modifying enzymes. Drug Resist.
623   Updat. 13, 151–171. https://doi.org/10.1016/j.drup.2010.08.003
624 Raza, S., Shin, H., Hur, H.-G., Unno, T., 2022. Higher abundance of core antimicrobial resistant
625   genes in effluent from wastewater treatment plants. Water Res. 208, 117882.
626   https://doi.org/10.1016/j.watres.2021.117882
627 Rose, A.N., Baggs, J., Wolford, H., Neuhauser, M.M., Srinivasan, A., Gundlapalli, A.V., Reddy,
628   S., Kompaniyets, L., Pennington, A.F., Grigg, C., Kabbani, S., 2021. Trends in antibiotic
629   use in United States hospitals during the Coronavirus Disease 2019 Pandemic. Open
630   Forum Infect Dis 8, ofab236. https://doi.org/10.1093/ofid/ofab236
631 Rothman, J., Sagir, A., Chung, S.-A., Boyajian, N., Dinh, T., Kim, J., Oval, J., Sharavanan, V.,
632   York, C., Zimmer-Faust, A., Langlois, K., Steele, J., Griffith, J., Whiteson, K., 2022.
633   Data for: Longitudinal metatranscriptomic sequencing of Southern California wastewater
634   representing 16 million people from August 2020-21 reveals widespread transcription of
635   antibiotic resistance genes. https://doi.org/10.7280/D11Q30
636 Rothman, J.A., Loveless, T.B., Griffith, M.L., Steele, J.A., Griffith, J.F., Whiteson, K.L., 2020.
637   Metagenomics of wastewater influent from Southern California wastewater treatment

638   facilities in the era of COVID-19. Microbiology Resource Announcements 9, 19–21.
639   https://doi.org/10.1128/mra.00907-20
640 Rothman, J.A., Loveless, T.B., Kapcia, J., 3rd, Adams, E.D., Steele, J.A., Zimmer-Faust, A.G.,
641   Langlois, K., Wanless, D., Griffith, M., Mao, L., Chokry, J., Griffith, J.F., Whiteson,
642   K.L., 2021. RNA viromics of Southern California wastewater and detection of SARS-
643   CoV-2 single-nucleotide variants. Appl. Environ. Microbiol. 87, e0144821.
644   https://doi.org/10.1128/AEM.01448-21
645 Schuierer, S., Carbone, W., Knehr, J., Petitjean, V., Fernandez, A., Sultan, M., Roma, G., 2017.
646   A comprehensive assessment of RNA-seq protocols for degraded and low-quantity
647   samples. BMC Genomics 18, 442. https://doi.org/10.1186/s12864-017-3827-y
648 Sims, N., Kasprzyk-Hordern, B., 2020. Future perspectives of wastewater-based epidemiology:
649   Monitoring infectious disease spread and resistance to the community level. Environ. Int.
650   139, 105689. https://doi.org/10.1016/j.envint.2020.105689
651 Slowikowski, K., 2018. ggrepel: Automatically position non-overlapping text labels with
652   "ggplot2." R package version 0. 8. 0.
653 Suzek, B.E., Wang, Y., Huang, H., McGarvey, P.B., Wu, C.H., UniProt Consortium, 2015.
654   UniRef clusters: a comprehensive and scalable alternative for improving sequence
655   similarity searches. Bioinformatics 31, 926–932.
656   https://doi.org/10.1093/bioinformatics/btu739
657 Symonds, E.M., Griffin, D.W., Breitbart, M., 2009. Eukaryotic viruses in wastewater samples
658   from the United States. Appl. Environ. Microbiol. 75, 1402–1409.
659   https://doi.org/10.1128/AEM.01899-08
660 Wickham, H., 2009. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York.
661 Wilder, M.L., Middleton, F., Larsen, D.A., Du, Q., Fenty, A., Zeng, T., Insaf, T., Kilaru, P.,
662   Collins, M., Kmush, B., Green, H.C., 2021. Co-quantification of crAssphage increases
663   confidence in wastewater-based epidemiology for SARS-CoV-2 in low prevalence areas.
664   Water Res X 11, 100100. https://doi.org/10.1016/j.wroa.2021.100100
665 Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2.
666   Genome Biol. 20, 257. https://doi.org/10.1186/s13059-019-1891-0
667 World Health Organization, 2021. Global antimicrobial resistance and use surveillance system
668   (GLASS) report: 2021. World Health Organization.
669 Wu, F., Zhang, J., Xiao, A., Gu, X., Lee, W.L., Armas, F., Kauffman, K., Hanage, W., Matus,
670   M., Ghaeli, N., Endo, N., Duvallet, C., Poyet, M., Moniz, K., Washburne, A.D.,
671   Erickson, T.B., Chai, P.R., Thompson, J., Alm, E.J., 2020. SARS-CoV-2 titers in
672   wastewater are higher than expected from clinically confirmed cases. mSystems 5,
673   e00614-20. https://doi.org/10.1128/mSystems.00614-20
674 Wu, L., Ning, D., Zhang, B., Li, Y., Zhang, P., Shan, X., Zhang, Q., Brown, M.R., Li, Z., Van
675   Nostrand, J.D., Ling, F., Xiao, N., Zhang, Y., Vierheilig, J., Wells, G.F., Yang, Y., Deng,
676   Y., Tu, Q., Wang, A., Global Water Microbiome Consortium, Zhang, T., He, Z., Keller,
677   J., Nielsen, P.H., Alvarez, P.J.J., Criddle, C.S., Wagner, M., Tiedje, J.M., He, Q., Curtis,
678   T.P., Stahl, D.A., Alvarez-Cohen, L., Rittmann, B.E., Wen, X., Zhou, J., 2019. Global
679   diversity and biogeography of bacterial communities in wastewater treatment plants. Nat
680   Microbiol 4, 1183–1195. https://doi.org/10.1038/s41564-019-0426-5
681 Xagoraraki, I., O'brien, E., 2020. Wastewater-based epidemiology for early detection of viral
682   outbreaks. Women in Water Quality 75. https://doi.org/10.1007/978-3-030-17819-2_5

683   Yang, Y., Li, B., Ju, F., Zhang, T., 2013. Exploring variation of antibiotic resistance genes in
684   activated sludge over a four-year period through a metagenomic approach. Environ. Sci.
685   Technol. 47, 10197–10205. https://doi.org/10.1021/es4017365
686   Yin, X., Yang, Y., Deng, Y., Huang, Y., Li, L., Chan, L.Y.L., Zhang, T., 2021. An assessment of
687   resistome and mobilome in wastewater treatment plants through temporal and spatial
688   metagenomic analysis. Water Res. 209, 117885.
689   https://doi.org/10.1016/j.watres.2021.117885
690   Zimmer-Faust, A.G., Steele, J.A., Xiong, X., Staley, C., Griffith, M., Sadowsky, M.J., Diaz, M.,
691   Griffith, J.F., 2021. A Combined digital PCR and next generation DNA-sequencing based
692   approach for tracking nearshore pollutant dynamics along the southwest United
693   States/Mexico border. Front. Microbiol. 12, 674214.
694   https://doi.org/10.3389/fmicb.2021.674214
695

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	4,000,000
Joint Water Pollution Control Plant	15	August 11 2020 – November 17 2020	400	4,800,000
North City Water Reclamation Plant	7	August 14 2020 – November 6 2020	30	1,400,000
Orange County Reclamation Plant #1	17	August 12 2020 – December 21 2020	140	2,600,000
Point Loma Water Treatment Plant	77	August 13 2020 – August 3 2021	175	2,200,000
San Jose Creek Water Reclamation Plant	15	August 12 2020 – November 18 2020	100	1,000,000
South Bay Water Reclamation Plant	7	August 13 2020 – November 5 2020	15	107,000

696

697   Table 1. Descriptions of the experiment sampling scheme and relevant information about each

698   WTP.

699   Figure legends:

700

701   Figure 1. Diagram indicating the date ranges of samples separated by wastewater treatment plan

702   and month. Y-axis codes correspond to abbreviated WTP names: ESC = Escondido Hale Avenue

703   Resource Recovery Facility, HTP = Hyperion Water Reclamation Plant, JWPCP = Joint Water

704   Pollution Control Plant, NC = North City Water Reclamation Plant, OC = Orange County

705   Reclamation Plant #1, PL = Point Loma Water Treatment Plant, SJ = San Jose Creek Water

706   Reclamation Plant, and SB = South Bay Water Reclamation Plant.

707

708   Figure 2. Stacked bar plots showing the relative abundances of RNA reads mapping to A)

709   unclassified taxonomic ranks, bacteria, viruses, and archaea; B) AMR genes separated by the ten

710   most abundant antibiotic classes each gene confers resistance to plus all others; C) ten most

711   abundant bacterial families plus all others; D) ten most abundant viral species plus all others. All

712   plots are faceted by WTP and labeled with sampling date.

713

714   Figure 3. Nonmetric multidimensional scaling ordination of Bray-Curtis dissimilarities of AMR

715   genes at greater than 0.01% relative abundance across A) all WTPs August – November 2020,

716   and B) ESC, HTP, and PL across August 2020 – August 2021. C) Heatmaps of the $\log_{10}$-

717   transformed counts of differentially abundant non-rRNA AMR genes across all WTPs August –

718   November 2020, and D) rRNA gene mutations conferring resistance to antimicrobials.

719   Hierarchal clustering of genes in each heatmap is through the Ward D2 algorithm. E) Bar plots

720   indicating the non-RNA AMR genes across ESC, HTP, and PL that changed over time and F)

721   AMR rRNA gene mutations. X-axes denote the linear model coefficient of each gene's

722   relationship to time.

723

724   Figure 4. Nonmetric multidimensional scaling ordination of Bray-Curtis dissimilarities of

725   bacterial species at greater than 0.01% relative abundance across A) all WTPs August –

726   November 2020, and B) ESC, HTP, and PL across August 2020 – August 2021. C) Heatmap of

727   the $\log_{10}$-transformed counts of differentially abundant bacterial genera at greater than 0.1%

728   relative abundance across all WTPs August – November 2020. D) Bar plots indicating the

729   bacterial genera across ESC, HTP, and PL that changed over time (only genera with a $P_{adj} <$

730   0.001 shown). X-axes denote the linear model coefficient of each genus's relationship to time.

731

732   Figure 5. Nonmetric multidimensional scaling ordination of Bray-Curtis dissimilarities of viruses

733   at greater than 0.01% relative abundance across A) all WTPs August – November 2020, and B)

734   ESC, HTP, and PL across August 2020 – August 2021. C) Heatmap of the $\log_{10}$-transformed

735   counts of differentially abundant viruses across all WTPs August – November 2020. D) Bar plots

736   indicating the viruses across ESC, HTP, and PL that changed over time. X-axes denote the linear

737   model coefficient of each virus's relationship to time.

738

739   Figure 6. Nonmetric multidimensional scaling ordination of Bray-Curtis dissimilarities of

740   metabolic pathway transcripts per million across A) all WTPs August – November 2020, and B)

741   ESC, HTP, and PL across August 2020 – August 2021. C) Bar plots indicating the metabolic

742   pathway at greater than 0.01% relative abundance across ESC, HTP, and PL that changed over

743   time. X-axes denote the linear model coefficient of each metabolic pathway's relationship to
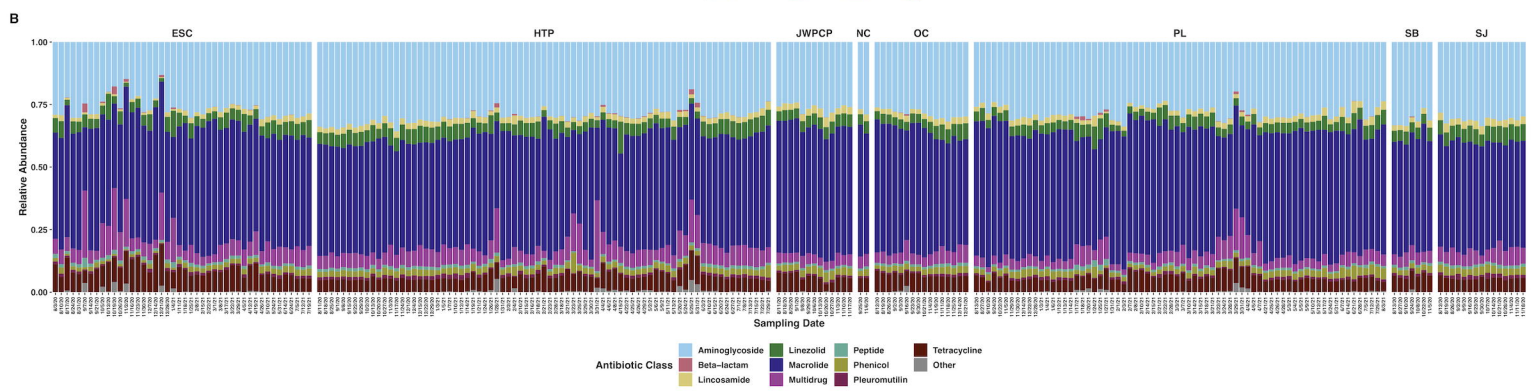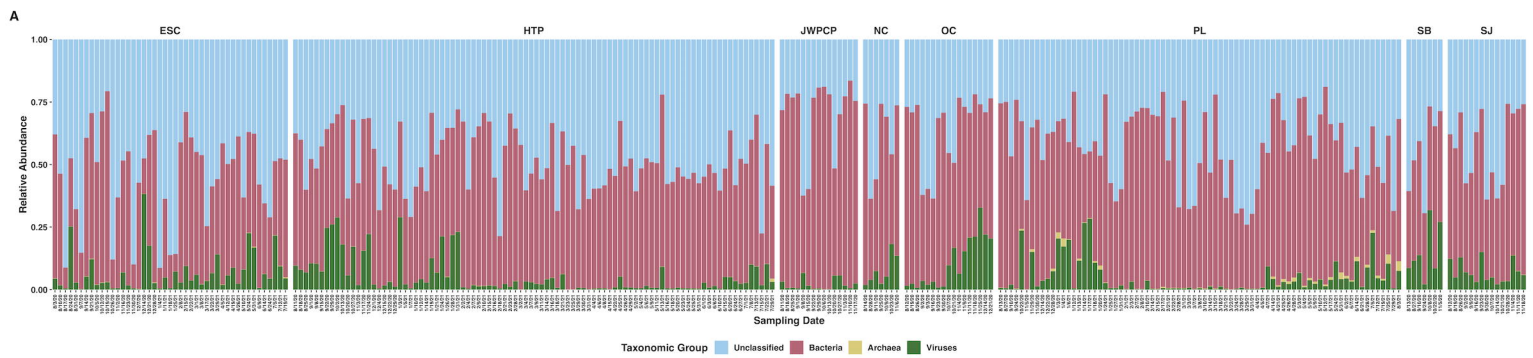
744   time.

745

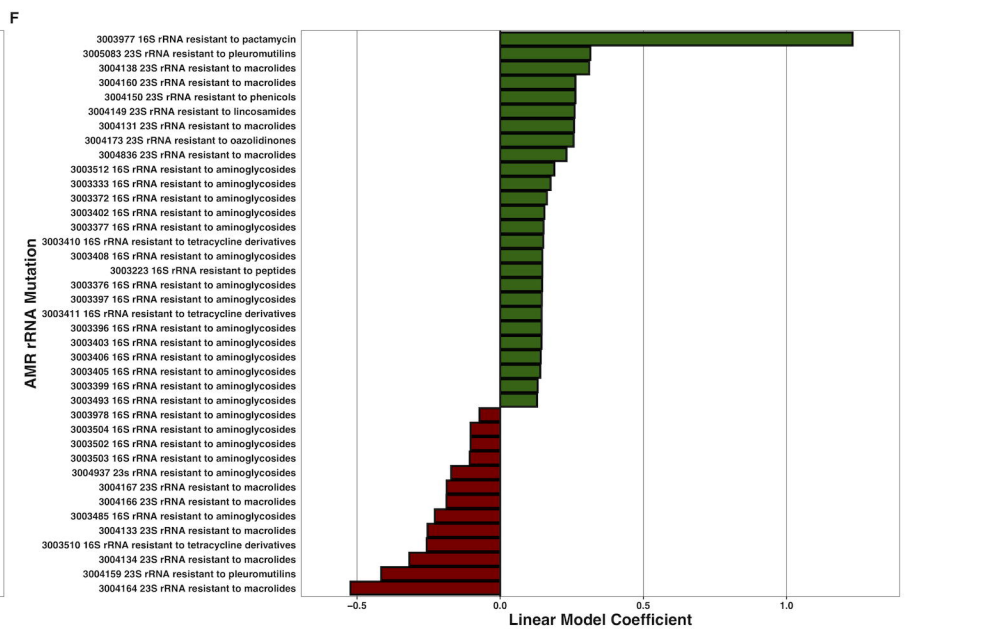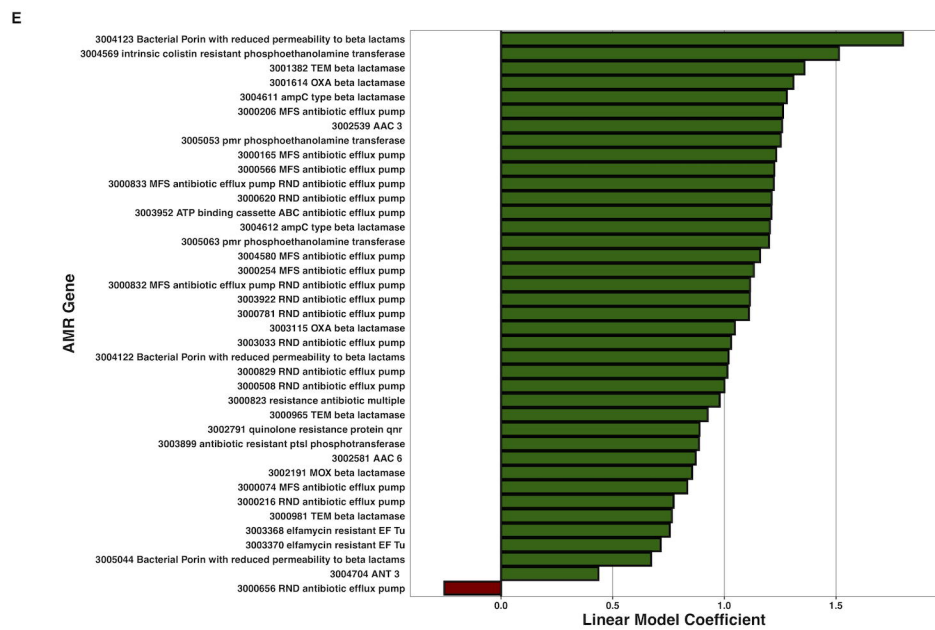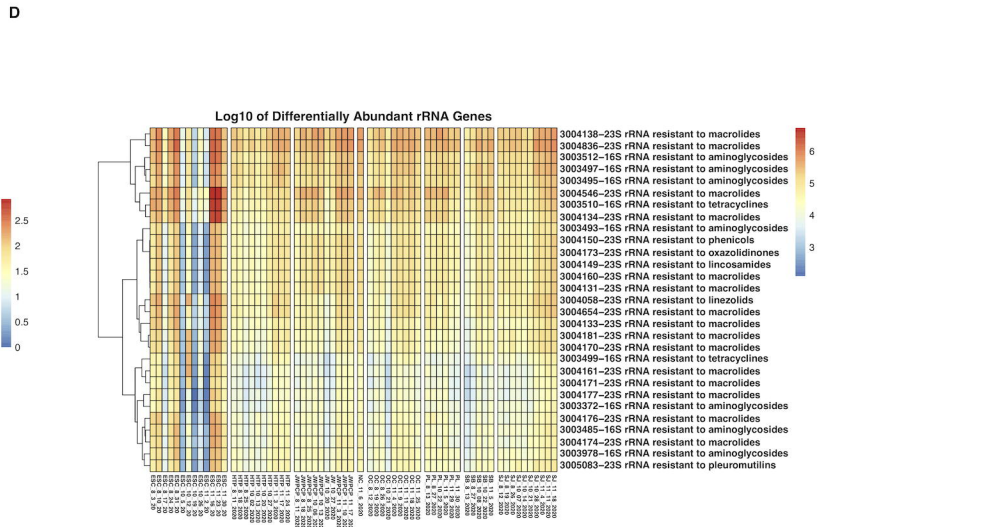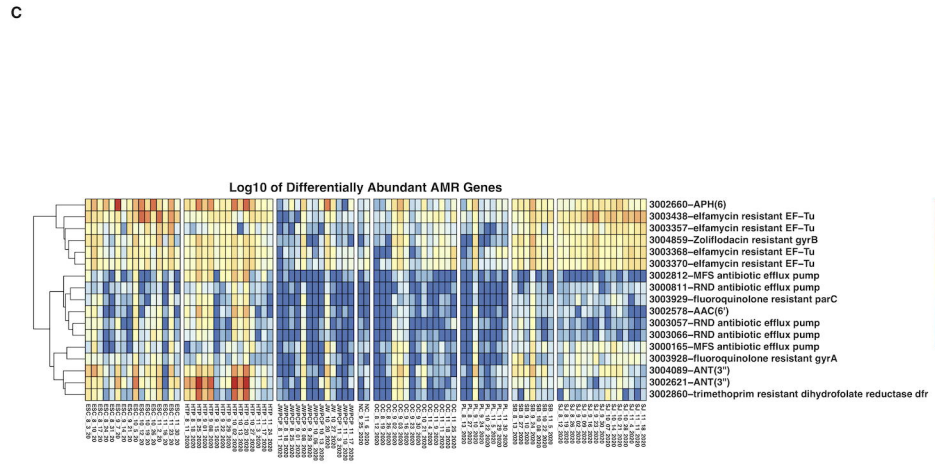746   Supplemental file legends:

747

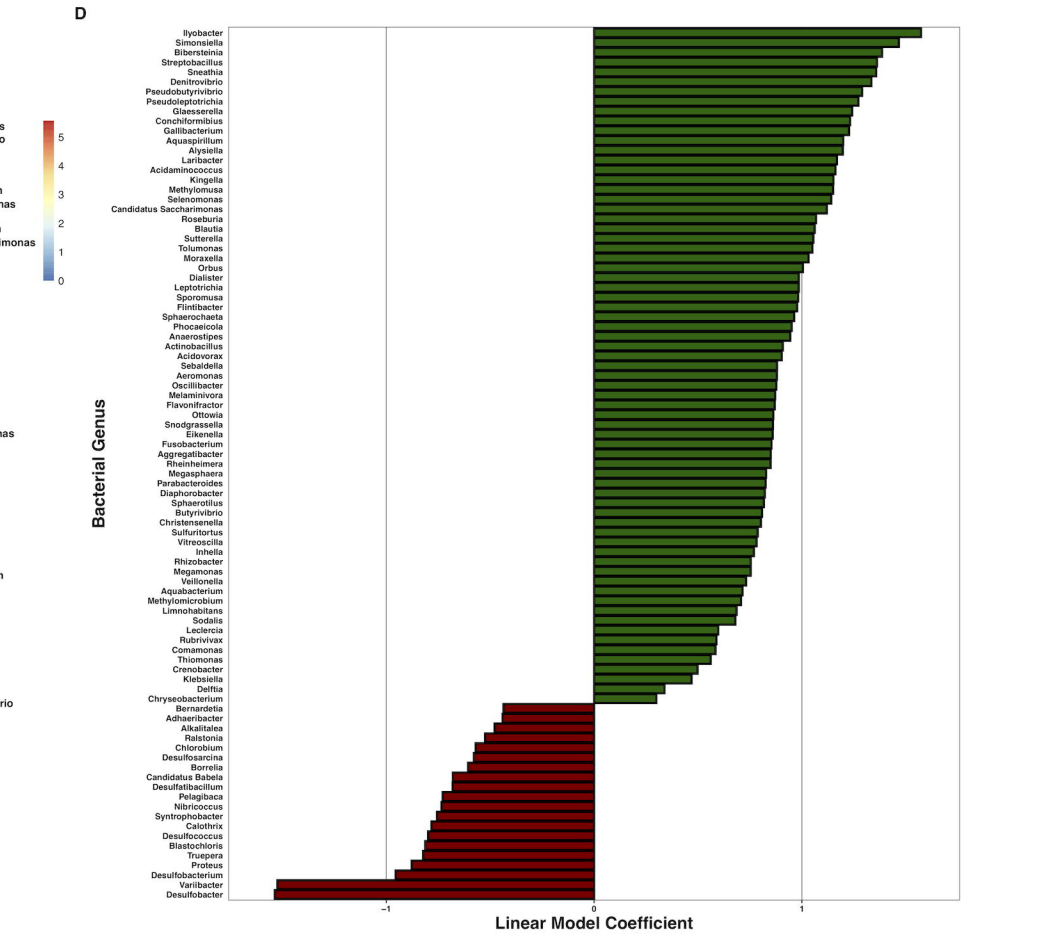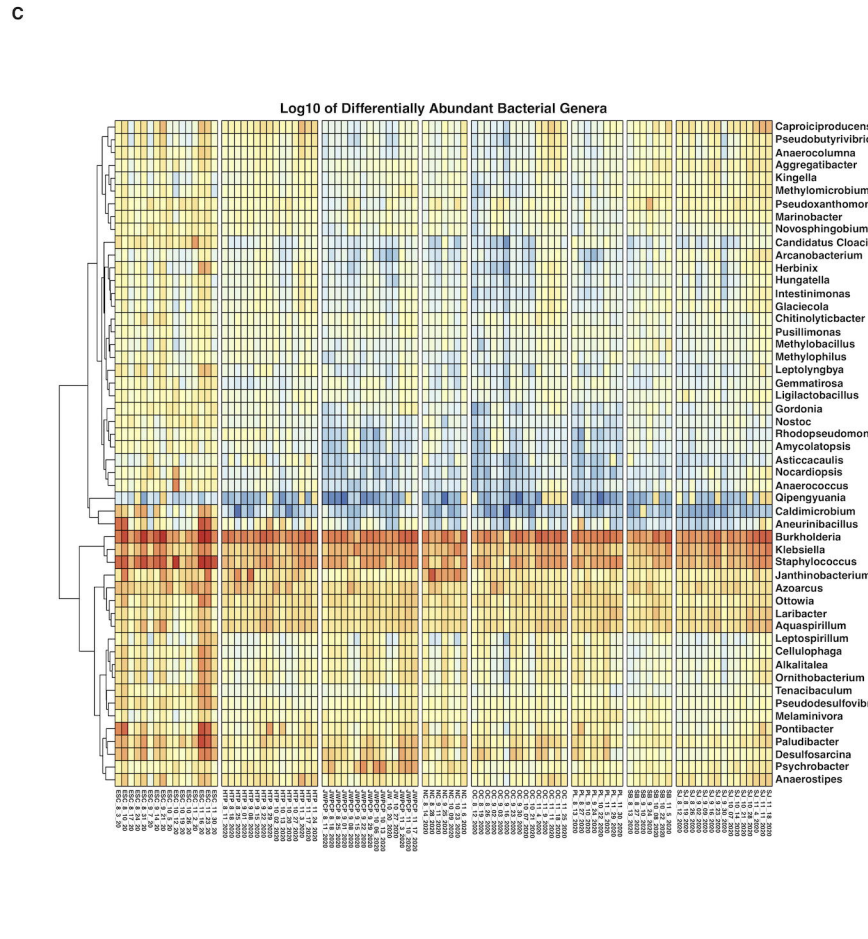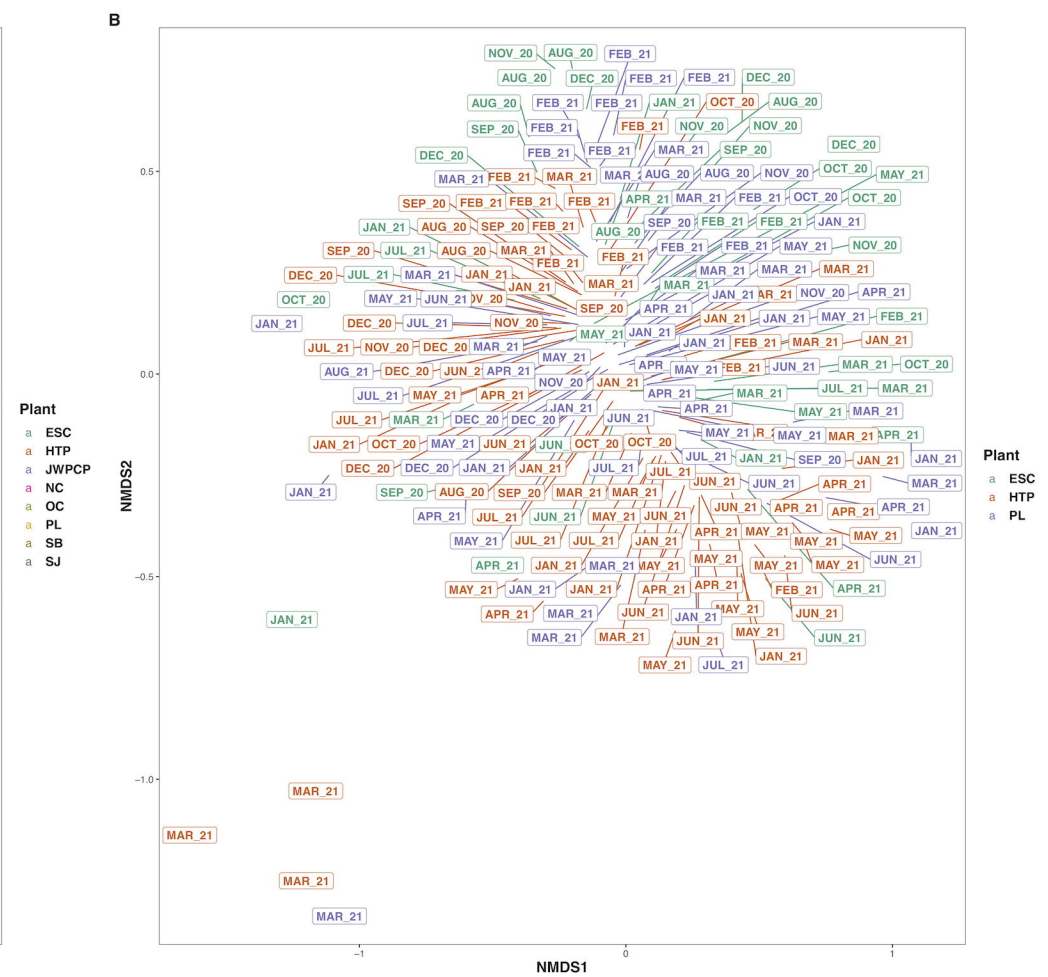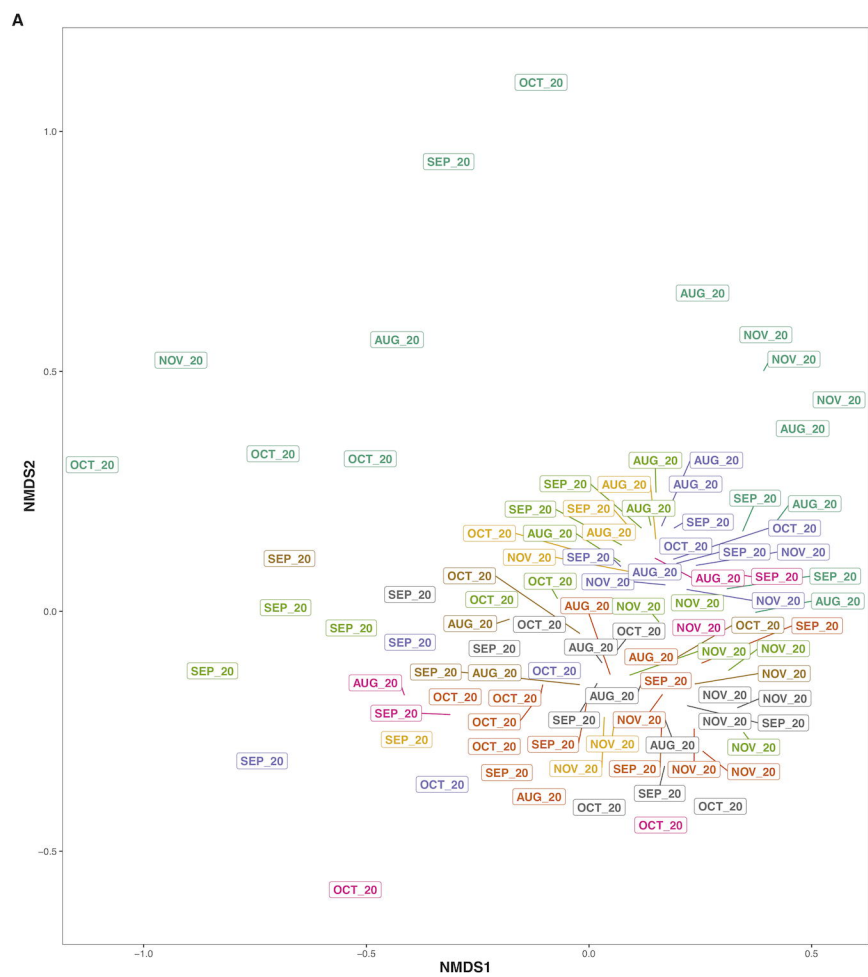748   Supplemental file SF1: Sample metadata.

749

750   Supplemental file SF2: ANCOM analyses outputs. Includes Wald scores, significance testing,

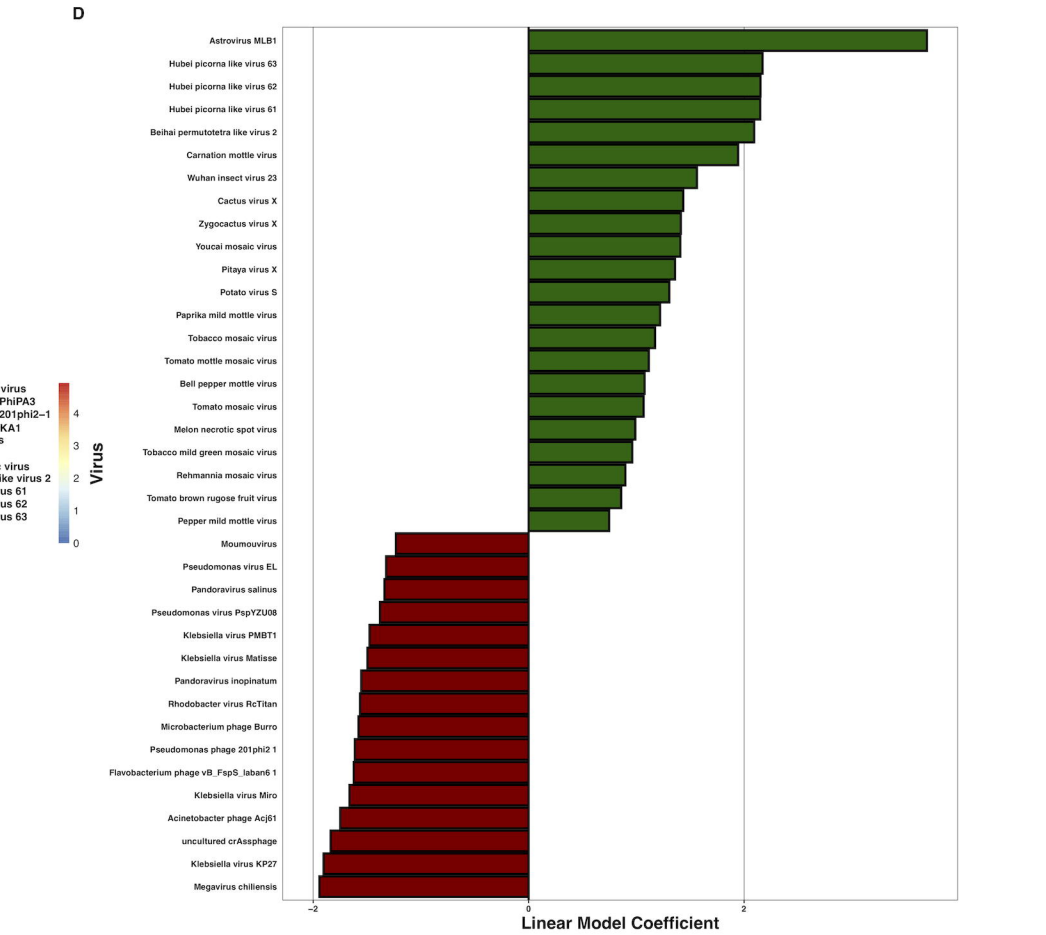751   and bacterial genus, virus, or ARO term being tested.
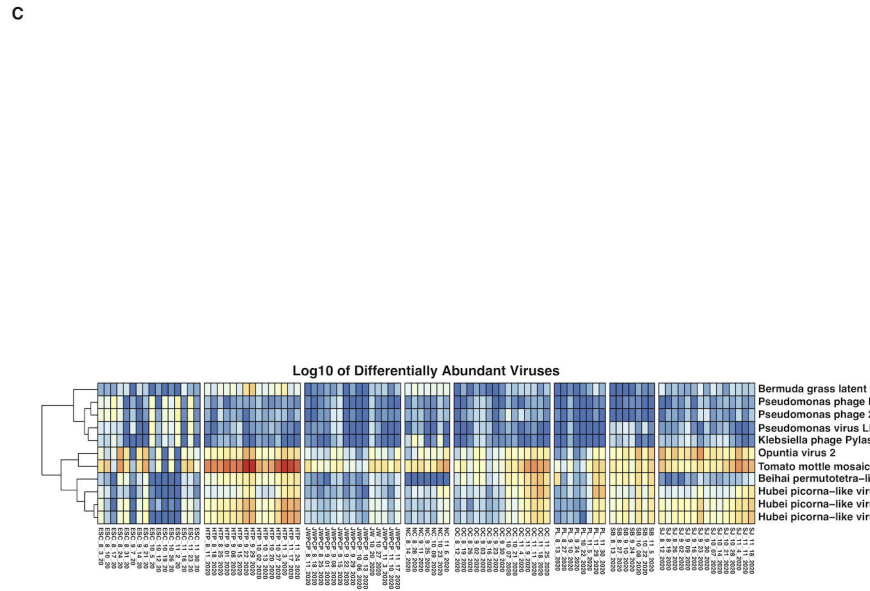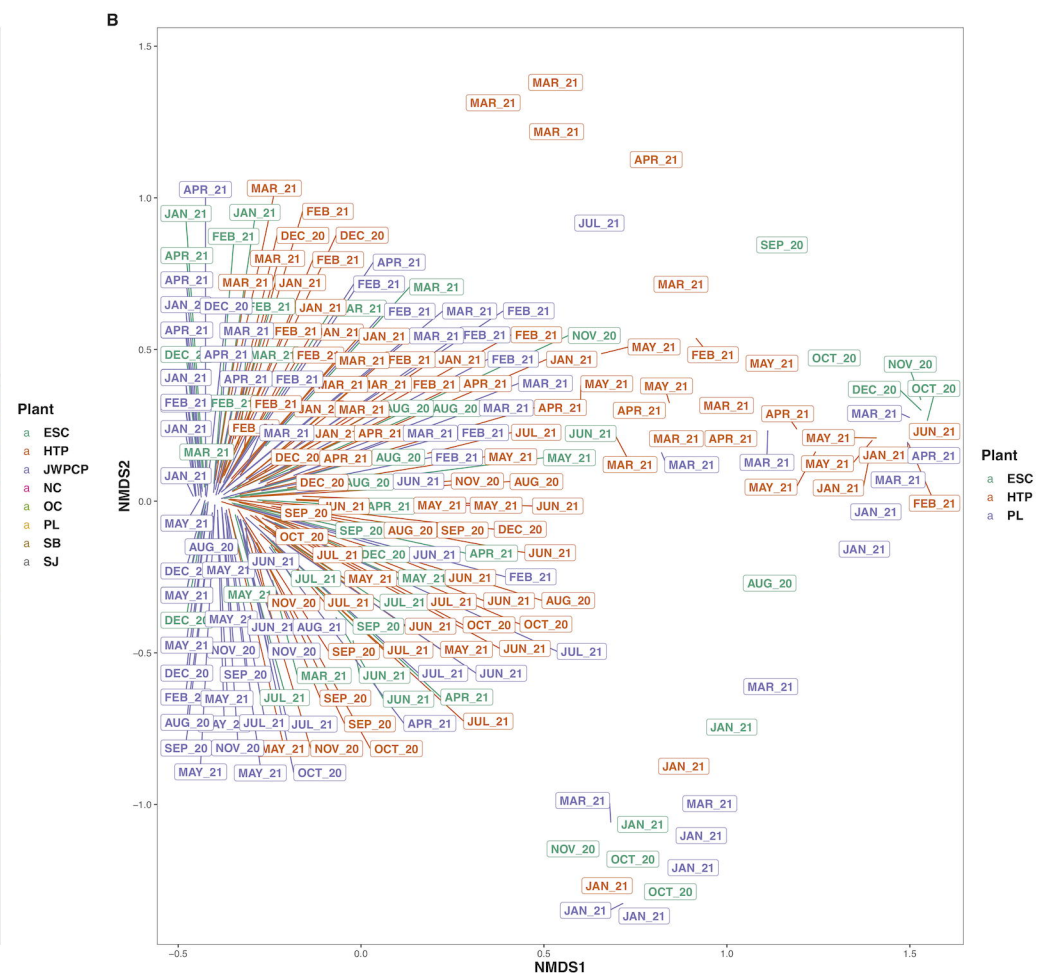
752

753   Supplemental file SF3: MaAsLin2 outputs. Includes linear model coefficients of proportional

754   abundances over time, standard errors, sample N included and excluded, P-value, and Q-value

755   for each bacterial genus, virus, ARO term, and HUMAnN3 pathway. Individual scatterplots for

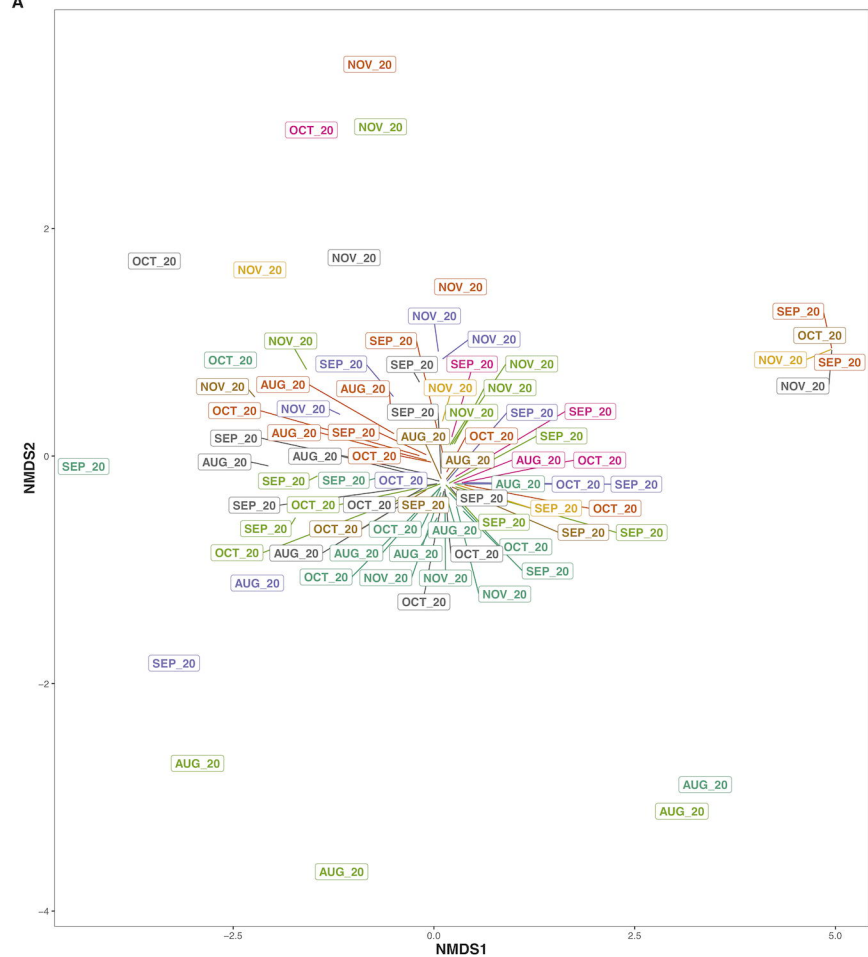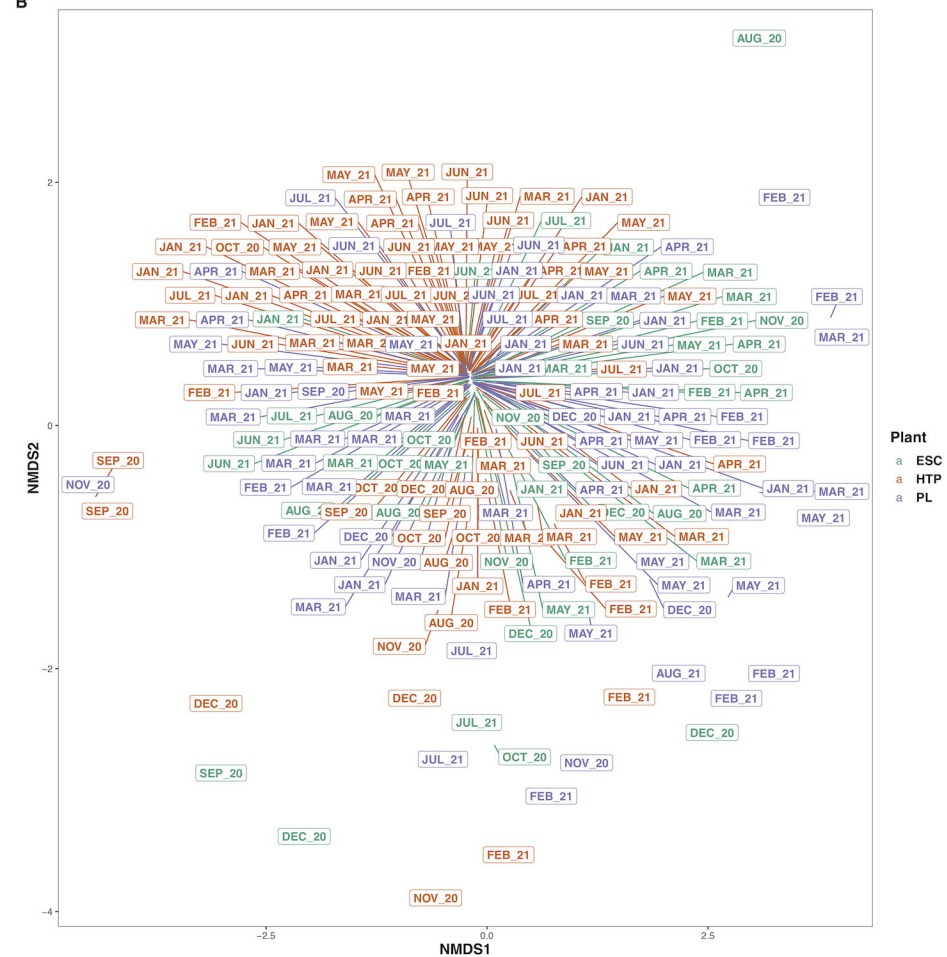756   each term being tested are available on Zenodo at (doi.org/10.5281/zenodo.6829029).

A

B

C Log10 of Differentially Abundant Viruses

D

Figure showing three panels. Panels A and B are NMDS ordination plots with NMDS1 on the x-axis and NMDS2 on the y-axis, with points labeled by month and year (e.g., NOV_20, OCT_20, SEP_20, AUG_20). Panel A legend "Plant" includes ESC, HTP, JWPCP, NC, OC, PL, SB, SJ. Panel B legend "Plant" includes ESC, HTP, PL. Panel C is a horizontal bar chart of Metabolic Pathway vs Linear Model Coefficient.

Panel C pathways (top to bottom):
- FAO PWY fatty acid beta oxidation I generic
- PWY 6282 palmitoleate biosynthesis I from 5Z dodec 5 enoate
- PWY 7664 oleate biosynthesis IV anaerobic
- PWY 5022 4–aminobutanoate degradation V
- PWY 7388 octanoyl acyl carrier protein biosynthesis
- FASYN ELONG PWY fatty acid elongation saturated
- DENITRIFICATION PWY nitrate reduction I denitrification
- PWY0 1296 purine ribonucleosides degradation
- GLYCOGENSYNTH PWY glycogen biosynthesis I from ADP D Glucose
- PWY 5189 tetrapyrrole biosynthesis II from glycine
- PWY 4984 urea cycle
- OANTIGEN PWY O antigen building blocks biosynthesis E coli
- PPGPPMET PWY ppGpp metabolism