

1 **Targeted metagenomic sequencing for detection of vertebrate viruses in wastewater for**
2 **public health surveillance**

3
4 ¹Camille McCall, ^{1,2}Ryan A. Leo Elworth, ³Kristine M. Wylie, ³Todd N. Wylie, ²Katherine
5 Dyson, ²Ryan Doughty, ²Todd J. Treangen, ^{4,5}Loren Hopkins, ⁵Katherine Ensor, ^{1,*}Lauren B.
6 Stadler

7
8 ¹ Department of Civil and Environmental Engineering, Rice University, Houston, TX, 77005

9 ² Department of Computer Science, Rice University, Houston, TX, 77005

10 ³ Department of Pediatrics, Washington University School of Medicine, St. Louis, Missouri
11 63110

12 ⁴ Houston Health Department, 8000 N. Stadium Dr., Houston, TX, 77054

13 ⁵ Department of Statistics, Rice University, Houston, TX, 77005

14
15 *Corresponding Author: Lauren Stadler

16 Mailing address:

17 6100 Main Street, MS-519

18 Houston, TX 77005

19 Email: lauren.stadler@rice.edu

20
21 **Abstract**

22
23 Viruses of concern for quantitative wastewater monitoring are usually selected as a result of an
24 outbreak and subsequent detection in wastewater. However, targeted metagenomics could
25 proactively identify viruses of concern when used as an initial screening tool. To evaluate the
26 utility of targeted metagenomics for wastewater screening, we used ViroCap, a panel of probes
27 designed to target all known vertebrate viruses. Untreated wastewater was collected from

28 wastewater treatment plants (WWTPs) and building-level manholes associated with vulnerable
29 populations in Houston, TX. We evaluated differences in vertebrate virus detection between
30 WWTP and building-level samples, classified human viruses in wastewater, and performed
31 phylogenetic analysis on astrovirus sequencing reads to evaluate targeted metagenomics for
32 subspecies level classification. Vertebrate viruses varied widely across building-level samples.
33 Rarely detected and abundant viruses were identified in WWTP and building-level samples,
34 including enteric, respiratory, and bloodborne viruses. Furthermore, full length genomes were
35 assembled from astrovirus reads and two human astrovirus serotypes were classified in wastewater
36 samples. This study demonstrates the utility of targeted metagenomics as an initial screening step
37 for public health surveillance.

38

39 **Keywords:** Wastewater-based epidemiology, metagenomic shotgun sequencing, viruses, capture
40 sequencing, astroviruses

41

42 **Synopsis:** This work demonstrates the utility of targeted metagenomic shotgun sequencing to
43 screen for human viruses in centralized and building-level wastewater samples.

44

45 **1. Introduction**

46

47 Approximately 75% of the US population is connected to a municipal wastewater treatment facility
48 (Westerhoff et al., 2015), which collects sewage containing a diverse range of commensal and
49 pathogenic organisms. Human vertebrate viruses excreted through feces, urine, skin, saliva, and
50 blood can be detected in wastewater to gather insights into viral diseases circulating within a

51 population. Previous studies have associated viral presence in wastewater with outbreaks in the
52 corresponding community (Bisseux et al., 2018; Farkas et al., 2018; Fernandez-Cassi et al., 2021;
53 Galani et al., 2022; Hellmér et al., 2014; McCall et al., 2021; Peccia et al., 2020; Prevost et al.,
54 2015; Zhao et al., 2022). More recently, the CDC and public health departments have deemed
55 wastewater monitoring as an integral public health surveillance tool for early detection and
56 mitigation of COVID-19 (Kirby et al., 2021). Viruses of concern for long-term quantitative
57 wastewater monitoring are usually selected as a result of clinical presence and subsequent
58 detection in municipal wastewater. Consequently, viruses that have significant incubation periods,
59 subclinical presentations, or are highly infectious may reach thousands of people before healthcare
60 facilities are notified and the virus is incorporated into the wastewater monitoring framework. The
61 integration of sensitive, unbiased methods into the existing framework is needed to proactively
62 identify and select viruses of concern for public health surveillance.

63
64 Untargeted metagenomic analysis provides a relatively unbiased approach to detecting viruses in
65 wastewater. Several studies have explored the presence of viral pathogens in wastewater using
66 metagenomics (Aw et al., 2014; Bibby and Peccia, 2013; Cantalupo et al., 2011; Fernandez-Cassi
67 et al., 2020; Martínez-Puchol et al., 2020; McCall et al., 2020; O'Brien et al., 2017). Metagenomics
68 has identified novel viruses, causative agents to disease outbreaks, and differences in viral diversity
69 between geographical regions (Matus et al., 2019; Newton et al., 2015; Ng et al., 2012; Xu et al.,
70 2011). Most wastewater viral metagenomic studies incorporate a sample purification step,
71 particularly, a concentration step, prior to processing samples for sequencing. Concentration is
72 important because human viruses are present at a low abundance in wastewater and wastewater
73 contains inhibitors that can interfere with the detection of such viruses. Membrane filtration is a

74 common concentration method used, either as a stand-alone method or in conjunction with other
75 methods, in wastewater metagenomic studies (McCall et al. 2019). Although concentration can
76 improve virus detection, it can impose bias during sample processing depending on several factors
77 such as the method used and virus morphology (Hjelmsø et al., 2017, Girón-Guzmán et al., 2023).
78 More studies are needed to explore the impact of concentration biases on virus detection in
79 metagenomic datasets.

80

81 More recently, targeted metagenomics has been introduced to increase the sensitivity of human
82 virus detection in wastewater. One such example of targeted metagenomic is in-solution
83 hybridization capture sequencing, which uses probes and magnetic bead technology to enrich for
84 a comprehensive set of vertebrate virus genomes (Wylie et al., 2015). This is especially important
85 for low abundance viruses such as human viruses that may go undetected in wastewater
86 metagenomes, which harbor a host of larger, more dominant viruses such as plant viruses and
87 bacteriophages. Furthermore, a previous study that compared targeted metagenomic sequencing
88 and amplicon deep sequencing was able to identify species-level information of several different
89 viruses, including enteroviruses, using both methods (Martínez-Puchol et al., 2020). This
90 highlights the potential for targeted metagenomic sequencing to identify species and subspecies of
91 important viral pathogens.

92

93 Wastewater sampling for metagenomic analysis is commonly carried out at a wastewater treatment
94 plant (WWTP). Samples collected from WWTPs can provide a snapshot of community health
95 across large populations, identify prevalent viruses circulating within a city or county, and provide
96 a means to accessible and convenient sampling locations. Although sampling at a WWTP is useful

97 for understanding population-wide disease burdens, viral disease burdens often vary between
98 different localized populations and demographic groups. A recent study found neighborhood-level
99 wastewater to contain biomarkers that are correlated with certain socio-economic and
100 sociodemographic factors (Choi et al., 2019). The pooling of wastewater in sewer networks can
101 create a homogenized view of viral presence, highlighting only viruses that are most abundant
102 across the population as a whole. Therefore, local outbreaks (e.g., neighborhood, building-level)
103 may go undetected in WWTP samples. Building-level sampling provides more geographically-
104 resolved data, higher specificity, and identification of viruses that are prevalent among high-risk
105 groups — specifically, children, elderly, incarcerated individuals, and the homeless. This
106 information can be used to select viruses for long-term quantitative wastewater monitoring
107 according to a community’s need and tailor intervention strategies to specific populations. Few
108 studies have implemented metagenomic sequencing for virus detection at the building-level
109 (Spurbeck et al., 2021) and to our knowledge no studies have implemented viral metagenomics on
110 the wastewater of vulnerable populations such as those inhabiting jails or homeless shelters.

111
112 This study evaluates targeted metagenomics via probe-based capture enrichment prior to
113 sequencing to facilitate the detection of vertebrate viruses for public health surveillance. Untreated
114 wastewater samples were collected from 7 WWTPs and 8 building-level manholes in Houston, TX
115 associated with high-risk populations, including a jail, homeless shelters, child care centers, a high
116 school, and nursing homes. Targeted and untargeted metagenomic datasets were compared to
117 quantify differences in virus detection between sequencing methods. Targeted and untargeted
118 samples underwent the same sample processing methods, specifically electronegative membrane
119 filtration (MF), prior to probe-based capture enrichment (i.e., targeted/MF and untargeted/MF).

120 Additionally, we compared vertebrate virus detection using targeted metagenomic shotgun
121 sequencing in samples concentrated using electronegative membrane filtration (targeted/MF) and
122 samples that underwent direct extraction (DE) (targeted/DE) to evaluate the impact of
123 concentration on virus detection. Of note, because this study focuses on the impact of enrichment
124 methods on virus detection in metagenomic data, untargeted/DE versus untargeted/MF and
125 targeted/DE versus untargeted/DE were not considered. Vertebrate virus detection was then
126 compared between WWTP and building-level wastewater samples for metagenomic datasets that
127 were concentrated and subsequently targeted using capture sequencing. A discussion of the human
128 viruses that were classified in all concentrated/targeted wastewater samples is also provided.
129 Species-level sequencing reads were further processed to assess the potential for targeted
130 metagenomics to identify viruses at the subspecies level in wastewater samples. This study focused
131 on astroviruses since they are abundant in wastewater and an important pathogen among children
132 and elderly.

133

134 **2. Methods**

135

136 **2.1. Study areas and wastewater sample collection**

137

138 Centralized wastewater samples were collected from the influent of 7 WWTPs (I-O) within
139 Houston, Texas covering 7 different sewersheds. In total, the WWTPs collect wastewater from
140 approximately 86% of the city's population. Building-level samples were collected from 8
141 manhole locations (A-H) with each one receiving untreated wastewater directly from their
142 respective facilities (child care center (A, B), high school (C), homeless shelter (F, G), jail (H),

143 nursing home (D, E)). All samples were collected from 24-hour composite samplers. Wastewater
144 samples were aliquoted into 250- or 500-mL sterile Nalgene bottles, and transported on ice to the
145 Rice University NEST lab for further processing.

146

147 **2.2. Sample concentration and extraction**

148

149 Samples undergoing filtration were concentrated within 24 hours of arrival. Concentration was
150 performed in duplicate for each wastewater sample by first aliquoting roughly 50 mL of sample
151 into 50mL conical tubes and centrifuging for 10 minutes at 4,100 g and 4°C to remove solids.
152 Samples were then filtered using electronegative membrane HA filters (HAWG047S6, Millipore
153 Sigma). Samples filters were transferred to 2 mL bead mill tubes followed by bead beating as
154 previously described (LaTurner et al., 2021) and briefly explained below. To evaluate the
155 difference between targeted and untargeted sequencing, and membrane filtration and direct
156 extraction on vertebrate virus detection in metagenomic datasets, WWTP samples M and O were
157 divided into two additional replicates each. One out of two of the additional replicates for each
158 WWTP sample were filtered as described above, and the remaining replicate was processed using
159 direct extraction. For direction extraction, each 50 mL replicate was well mixed and 300 uL of
160 sample was transferred to a bead tube. For both filtered and unfiltered samples, bead beating tubes
161 contained 0.1 mm glass bead and 1 mL of lysis buffer. Bead beating was carried out with XX
162 instrument and YY conditions, and total nucleic acid was extracted on a chemagic 360 automated
163 platform using the Viral DNA/RNA 300 Kit H96 (CMG-1433, PerkinElmer) following the
164 manufacturer's protocol.

165

166 **2.3. Two-strand cDNA synthesis and probe-based capture sequencing**

167

168 To capture both DNA and RNA viruses, total nucleic acid from each sample was subject to double
169 stranded cDNA synthesis according to (Wang et al., 2003) with some modifications. Specifically,
170 the random primer was modified to contain only the nanomer region of the sequence to eliminate
171 the presence of Primer B nucleotides in sequencing reads. Two-strand cDNA synthesis for each
172 sample was carried out in a 30 uL reaction mixture containing 40 uM of random primer and 5 uL
173 of nucleic acid extract. The complete protocol, including reaction mixture components and
174 thermocycling conditions, is detailed in Appendix 1 of the supplementary information. Following
175 cDNA synthesis, samples were shipped on dry ice to the Genome Technology Access Center at
176 McDonnell Genome Institute at Washington University in Saint Louis, MO for indexed
177 sequencing library construction with the Kapa HyperPrep kit (Roche). For libraries that were
178 enriched with ViroCap, sequencing libraries were pooled and combined with the ViroCap panel,
179 which consists of probes designed to target all known DNA and RNA vertebrate viruses (Wylie et
180 al., 2015). Enrichment was carried out according to the manufacturer's instructions (Roche). The
181 two additional replicates that were filtered for WWTPs M and O were sequencing directly without
182 ViroCap enrichment. Illumina sequencing for targeted and untargeted samples was performed on
183 a NovaSeq platform generating 150 bp paired-end reads. Sequences were demultiplexed using the
184 index sequences from each sample's sequencing library. The raw sequence reads used in this study
185 can be found in the Sequence Read Archive (SRA) under BioProject ID: PRJNA943189.

186

187 **2.4. Metagenomic and phylogenetic analysis**

188

189 Sequencing reads were analyzed using the ViroMatch pipeline (Wylie and Wylie, 2021). In
190 summary, reads were quality filtered, host screened, and aligned to virus-specific reference
191 genomes obtained from NCBI GenBank. Multistep alignment was performed using BWA-MEM
192 for nucleotide mapping and DIAMOND for translated nucleotide mapping. If a virus obtained
193 assigned reads for both alignment strategies, the assigned reads were combined to get the total
194 number of assigned reads for that virus. A threshold of 0.1% of the total assigned reads across all
195 samples for each virus detected was subtracted as background to account for sequencing crosstalk
196 observed in pooled sequencing libraries (Wylie et al., 2018).

197
198 Reads aligning to mamastrovirus (MAstV) genus were collected and further processed to assess
199 the genetic diversity of human astroviruses in wastewater samples. MAstV reads per sample were
200 combined into individual R1 and R2 FASTQ files based on their species-level classification and
201 assembled into contigs using IDBA-UD. IDBA-UD is useful for strain-level analysis due to its
202 high accuracy and genome recovery (Sutton et al., 2019). Contigs with sequencing lengths
203 comparable to the complete genome of the appropriate MAstV were considered. Putative whole
204 genome contigs were aligned against their respective sequencing reads to determine the percent
205 coverage and average read depth per contig using BWA-MEM under default conditions. Contig
206 alignment with BWA-MEM was visually inspected using the Integrative Genomics Viewer (IGV)
207 to evaluate the quality of each contig. Potential misassembled contigs were amended, if feasible,
208 before inclusion in multiple sequence alignment (MSA).

209
210 In our samples, MAstV-1 was the only astrovirus species for which we obtained contigs
211 comparable to the length of its whole genome (~6.7 kbp). Sequence assembly generated a total of

212 3 MAsV-1 contigs from child care center A (n=2) and WWTP M (n=1). Child care center A-1
213 contig and the WWTP M contig were divided at 5.8 kbp and 4.4 kbp, respectively, to remove gaps
214 in read coverage. Therefore, child care center A-1 and WWTP M contigs were considered near
215 complete genome contigs. Contig coverage results and lengths are provided in Table S1 and
216 Figures S1-S3.

217
218 Homology and phylogenetic analyses were carried out with 38 complete MAsV reference
219 genomes obtained from NCBI GenBank, representing each human MAsV-1 serotype, and used
220 as references for MSA and phylogenetic analysis. MSA was performed using MUSCLE. All
221 alignment files were forwarded to Fasttree 2.1.11 to create maximum likelihood trees using the
222 generalized time-reversible model and a bootstrap test with 1000 replicates. The remainder of the
223 parameters were run under default conditions. Tree files were generated in Newick format and
224 visualized using MEGA 11.0.

225

226 **2.5. Statistical Analysis**

227

228 All statistical analyses were performed in R (R Core Team, 2022). Prcomp was used for the
229 PCoA. Welch's student t-test was used to evaluate the difference between samples after verifying
230 the normality of the datasets using the Shapiro-Wilk test and qq plots. A p-value < 0.05 was
231 considered statistically significant.

232

233 **3. Results and Discussion**

234

235 Metagenomic shotgun sequencing enriched by ViroCap probe-based hybridization capture was
236 used to screen WWTP and building-level wastewater for vertebrate viruses including human
237 pathogens. Targeted sequencing was performed on 8 building-level samples (A-H) and 7 WWTP
238 samples (I-O). Illumina sequencing resulted in a total of 171 Gb (gigabases) or approximately 1.2
239 billion reads. A total of 1.1 million (0.1%) reads were assigned to viral taxonomic groups via the
240 ViroMatch pipeline. After applying the contamination threshold, the number of viral affiliated
241 reads reduced by 0.09%.

242

243 **3.1. Targeted metagenomics with electronegative membrane filtration increased**
244 **vertebrate virus detection in wastewater samples when compared to untargeted and**
245 **direction extraction methods**

246

247 Wastewater samples M and O were divided into two additional replicates each to evaluate the
248 impact of electronegative membrane filtration and targeted sequencing on vertebrate virus
249 detection in sequencing reads. Targeted metagenomic sequencing coupled with membrane
250 filtration (targeted/MF) resulted in a 35% (n=20) and 67% (n=18) increase in the number of unique
251 vertebrate viral families detected in the M and O WWTPs, respectively, as compared to the
252 untargeted/MF samples for WWTPs M (n=13) and O (n=6) (Figure S4). All viral families that
253 were detected in the untargeted/MF sample were also detected in the targeted/MF sample for both
254 WWTPs. The number of reads per vertebrate viral family generally increased in targeted samples
255 with the exception of Poxviridae, which decreased in both wastewater samples (Table S2).
256 Poxviridae family contains viruses that infect vertebrates as well as viruses that infect Arthropods,
257 invertebrate animals. Hence, the decrease in Poxviridae reads was due to a decrease in viruses with

258 invertebrate hosts (data not shown), which are not included in the ViroCap panel to date.
259 Additional viral targets can be added to the panel if invertebrate viruses are of interest. Findings
260 here suggest that targeted metagenomic sequencing with ViroCap increases detection and read
261 counts of targeted viruses. This was consistent with other studies that used targeted metagenomics
262 for virus detection (Martínez-Puchol et al., 2020; Paskey et al., 2019; Wylie et al., 2015).

263
264 Similarly, targeted/MF increased the number of vertebrate viral families detected by more than
265 half in both M and O WWTP samples as compared to samples that were targeted and directly
266 extracted (targeted/DE), 80% (n= 4) and 67% (n=6), respectively (Figure S4). Similarly,
267 Segelhurst et al., 2022 used a tangential-flow filtration method to recover 53% more complete or
268 near complete SARS-CoV-2 genomes from wastewater samples as compared to concentration
269 using precipitation. Still, a previous study by Hjelmso et al., 2017 evaluated the impact of
270 polyethylene glycol precipitation (PEG), skimmed milk flocculation, glass wool filtration, and
271 monolithic adsorption filtration on the sequencing of viruses in wastewater. They demonstrated
272 that concentration methods significantly impacted the detection of viruses in metagenomic
273 sequencing reads with PEG concentrated samples obtaining more viral-affiliated reads compared
274 to the other methods. Although the application of the two above mentioned studies is different, it
275 is clear that more work is needed to understand the impact of concentration methods on the
276 presence of viruses in sequencing data. Concentration methods can also impose bias on viral
277 recovery considering the multiple forms in which viruses can exist in wastewater (i.e., intact, non-
278 intact) and differences in virus morphology (i.e., enveloped, nonenveloped) (Mondal et al., 2021).
279 In fact, Giron-Guzman et al. 2023 found that the Enviro Wastewater TNA concentration kit was
280 more effective at recovering non-enveloped viruses from wastewater as compared to an aluminum-

281 based adsorption-precipitation method. The opposite was true for enveloped viruses (Girón-
282 Guzmán et al., 2023). This makes direct extraction a plausible approach to circumvent such
283 challenges. However, raw wastewater contains high concentrations of organic matter and other
284 inhibitors that can interfere with the detection and sensitivity of later processes. This necessitates
285 the addition of one or multiple purification steps to increase the chances of virus detection in these
286 complex sample matrices. Additionally, since viruses are present at dilute concentrations in
287 wastewater, direct extraction can result in a lower equivalent volume of viruses in the final nucleic
288 acid extract. A previous study found the use of centrifugation, filtration and nuclease-treatment to
289 yield the highest proportion of viral affiliated sequences as compared to using only one or two of
290 the aforementioned methods (Hall et al., 2014). Further, LaTurner et al., 2021 demonstrated that
291 HA filtration with bead beating yielded the highest concentrations of SARS-CoV-2 as compared
292 to other concentration methods tested including direct extraction. In this study, the highest viral
293 yield was obtained when pairing membrane filtration with targeted metagenomic sequencing using
294 ViroCap. This combination of methods was therefore used for the subsequent analysis below.

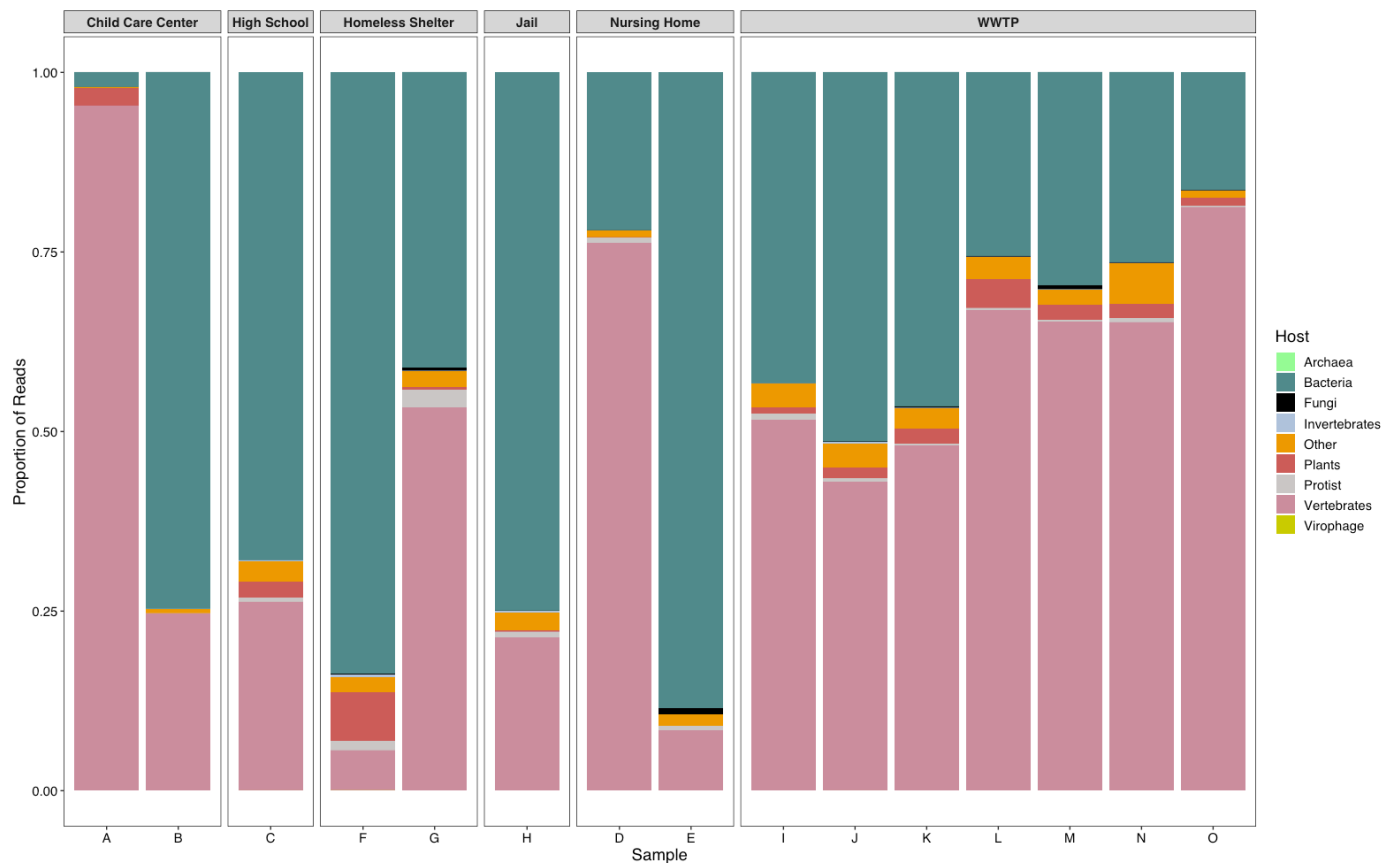
295

296 **3.2. Vertebrate virus profiles were more homogenous across WWTP as compared to** 297 **building-level samples**

298

299 Viruses associated with vertebrates and bacteria constituted the greatest proportion of reads across
300 all base samples (Figure 1). The total number of viral affiliated reads per sample is listed in Table
301 S4. The proportion of vertebrate viruses across wastewater samples range between 0.06–0.95 with
302 an average of 0.49. The average was approximately 26–43% higher than vertebrate virus detection
303 in untargeted metagenomic studies (Aw et al., 2014; Ng et al., 2012; O’Brien et al., 2017). The

304 proportion of vertebrate viruses was relatively consistent across WWTPs as compared to building-
305 level samples where host range noticeably varied. This fluctuation in virus hosts could be due to
306 differences in wastewater composition, diurnal cycles, and/or human input. Building-level samples
307 can be more sensitive to minor changes in community activity as compared to WWTP samples.
308



309
310 Figure 1. Proportion of hosts detected in WWTP (I-O) and building-level samples (A-H).

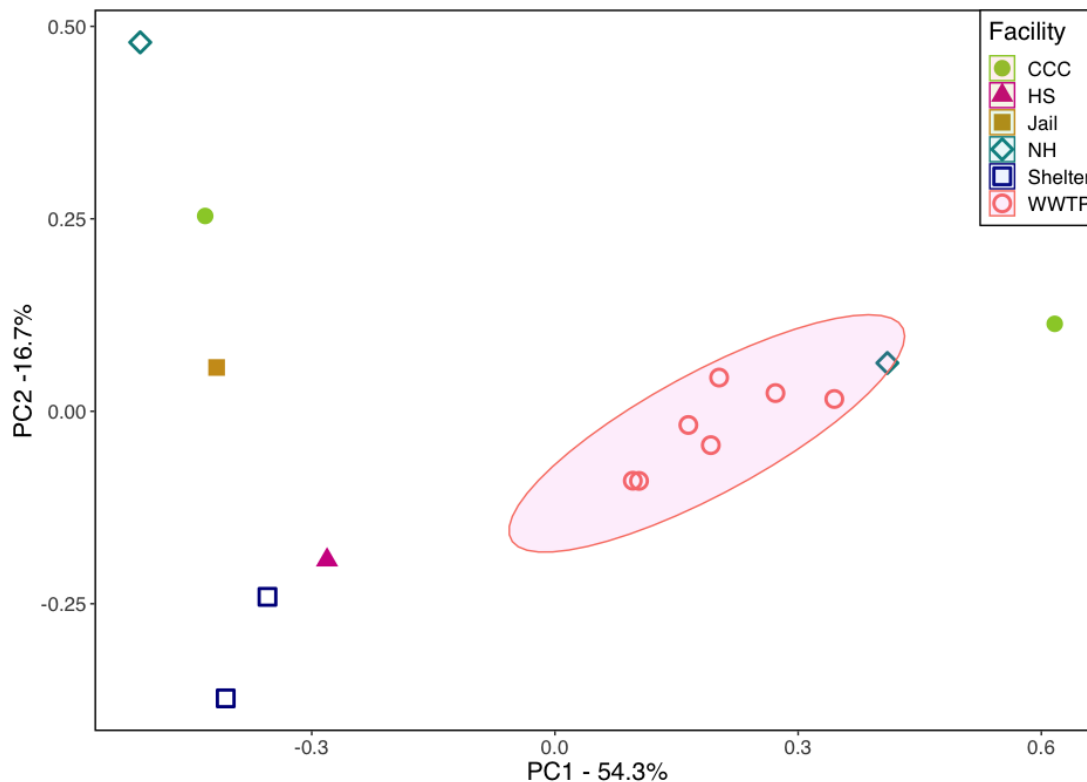
311
312 Twenty-four different vertebrate viral families were detected in wastewater samples with
313 *Astroviridae* followed by *Adenoviridae*, and *Parvoviridae* obtaining the greatest number of hits
314 across WWTPs. In total, *Astroviridae*, *Caliciviridae*, and *Picornaviridae* obtained the greatest
315 number of hits in building-level samples (Figure 2). A principal component analysis was

316 performed to determine the similarity between samples at the family taxonomic level (Figure 3).
 317 *Astroviridae*, *Coronaviridae*, and *Anelloviridae* viral families had the most influence when
 318 differentiating between samples. As expected, there were more similarities between WWTPs than
 319 building-level samples. The two nursing home samples were more similar to each other as
 320 compared to the pair of child care center and homeless shelter samples. More data is needed to
 321 sufficiently understand the differences in vertebrate virus detection between various communities.
 322



323
 324 Figure 2. Classification of vertebrate virus families detected in WWTP and building-level
 325 samples.

326

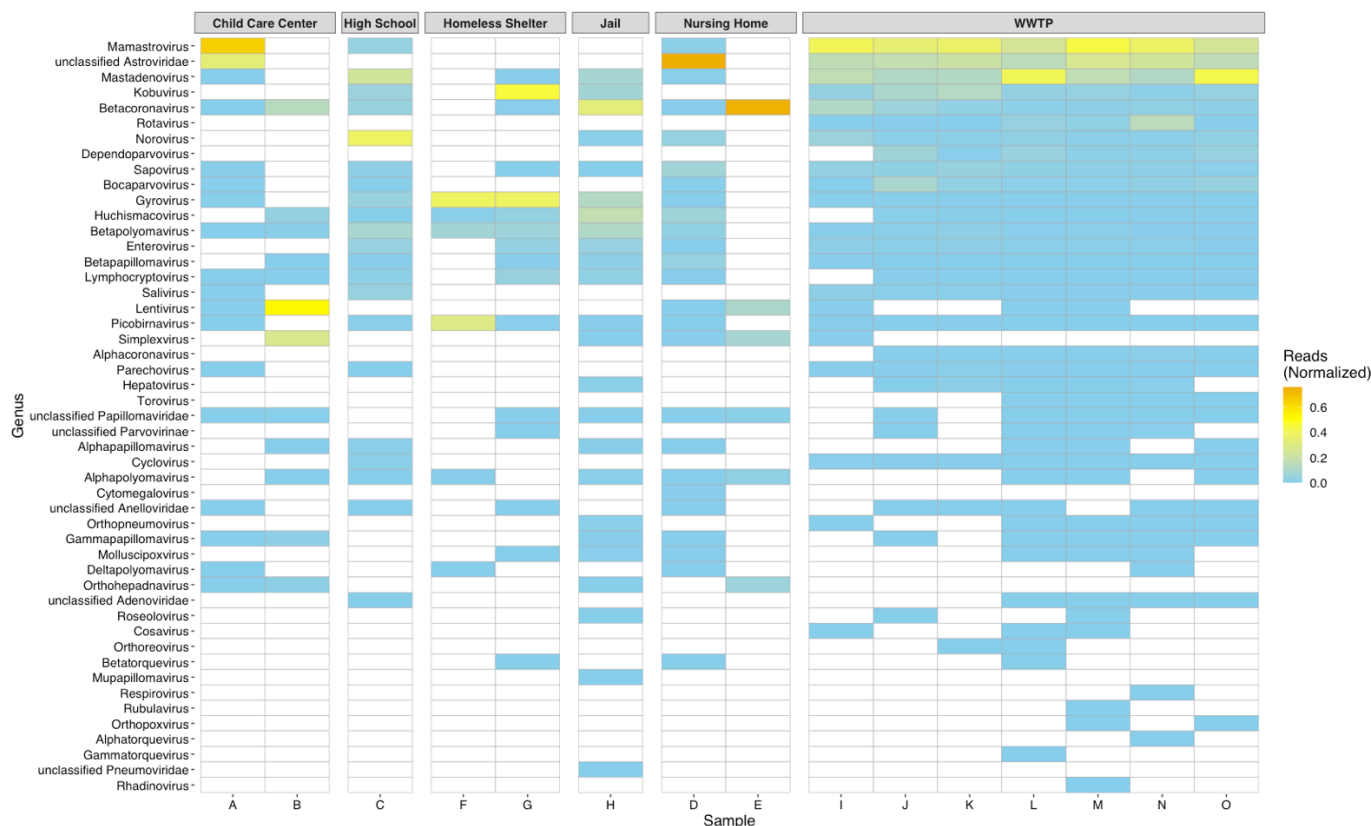


327
328 Figure 3. Principal component analysis (PCoA) of vertebrate viruses between sample types. Plot
329 is generated at the family level. The wastewater samples are abbreviated as follows: child care
330 center (CCC), high school (HS), jail (Jail), nursing home (NH), homeless shelter (Shelter), and
331 wastewater treatment plant (WWTP). The polygon represents the 95% confidence level for a
332 multivariate t-distribution.

333 334 **3.3. Classification of human vertebrate viruses in Houston wastewater**

335
336 Of the 24 vertebrate families detected, 20 of them were identified as human viral families
337 containing 43 genera and 6 unclassified viral categories. Targeted metagenomic analysis identified
338 respiratory, enteric, and other viral pathogens associated with blood, skin, and mucosal
339 transmissions (Figure 4, Table S5). There were important distinctions in virus communities

340 between WWTP and building-level samples. Generally, human viruses detected in building-level
341 samples were less homogenous than WWTPs samples. Only 19–30% of viruses detected in the
342 child care center, homeless shelter, or nursing home samples were also detected in the
343 corresponding sample belonging to the same building type. Of the 49 viral groups identified, 4
344 were detected in building-level samples that went undetected in WWTP samples. Interestingly, the
345 4 viral groups that were not detected in the WWTP samples are non-enteric viruses, with
346 orthohepadnavirus being rarely detected in wastewater samples (Table S5). Factors influencing
347 virus detection in centralized WWTPs include their shedding patterns in human excrements,
348 persistence in wastewater, dilution, and city-wide prevalence, as compared to localized
349 occurrences. Since enteric viruses tend to dominate wastewater samples, non-enteric viruses can
350 be challenging to detect, especially in large-scale WWTP samples. Findings here suggest that
351 sampling at the building-level can increase the potential to identify viruses less commonly found
352 in wastewater. Conversely, there were 12 viral groups detected in the WWTP samples that were
353 not present in building-level samples. These viruses include both non-enteric and enteric viruses
354 (Table S5). Since the WWTP samples contain wastewater collected from approximately 86% of
355 the city’s population, it is likely that other populations within the sewershed contributed to the
356 signal observed.



357
 358 Figure 4. Heatmap of human genus virus diversity and normalized abundance in each sample.
 359 The normalized abundance was calculated by dividing the number of reads assigned to each
 360 genus by the total number of human viral-assigned reads in the corresponding samples. White
 361 cells indicate absence of associated virus in the sample. Genera are in descending order
 362 according to abundance.

363
 364 Enteric followed by respiratory viruses collectively constituted the greatest proportion of reads in
 365 wastewater samples. Enteric pathogens generally cause gastrointestinal illnesses and are shed in
 366 significant concentrations in stool making them abundant in wastewater (Haramoto et al., 2018).
 367 Mamastrovirus (MAstV) and unclassified *Astroviridae* obtained the greatest number of hits among
 368 human viral reads. MAstV is a genus in the *Astroviridae* family infecting mammals. There are
 369 currently 19 species in the MAstV genus (MAstV-1-19) that are recognized by the International

370 Committee on Taxonomy of Viruses (ICTV). Human viruses are classified within the following
371 species: MAstV-1 (HAstV-1-8), MAstV-6 (MLB1-3), MAstV-8 (VA2/HMO-A, VA4), and
372 MAstV-9 (VA1/HMO-C, VA3/HMO-B). HAstV-1-8 are considered classic HAstVs while the
373 remaining serotypes are considered novel HAstVs (Donato and Vijaykrishna, 2017). In this study,
374 MAstV-1, -6, and unclassified HAstVs obtained the majority of HAstV reads. Astroviruses were
375 detected in 7 of 7 and 3 of 8 WWTPs and building-level samples, respectively, with child care
376 center A and nursing home D obtaining the greatest relative abundance among samples.

377
378 Astroviruses cause a significant burden mainly among pediatric and elderly populations (Vu et al.,
379 2017). They are a leading cause of gastroenteritis in children (behind rotavirus and norovirus) and
380 could account for 10% of all gastroenteritis outbreaks in child care centers in the U.S. (Lyman et
381 al., 2009). Similarly, elderly populations are at higher risk of gastroenteritis caused by bacterial
382 and viral infections, and individuals in nursing homes face an increased risk of mortality from such
383 infections mainly due to dehydration and comorbidities (Montoya and Mody, 2011).

384
385 Further, nursing home E obtained the greatest number of betacoronavirus reads. Severe acute
386 respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of coronavirus disease
387 2019 (COVID-19), was the most abundant species detected in the betacoronavirus genus (data not
388 shown). Since the start of the COVID-19 pandemic, wastewater monitoring of SARS-CoV-2 has
389 played a critical role in early detection of outbreaks and rapid implementation of public health
390 measures (Barrios et al., 2021; Hopkins et al., 2023; Wurtzer et al., 2022).

391

392 Interestingly, reads associated with lentivirus were detected in both child care centers and were
393 most abundant in child care center B. The lentivirus genus contains the human immunodeficiency
394 virus (HIV), which is the only species of lentiviruses that affects humans. Targeted metagenomics
395 identified reads homologous to HIV in several wastewater samples including both child care center
396 samples, suggesting the presence of the virus in wastewater and within the community. Perinatal
397 transmission is the most common route among infected children (Deeks et al., 2015). Although
398 newly infected HIV cases in children have declined since its initiation in 1970, HIV remains an
399 ongoing pandemic (Deeks et al., 2015).

400
401 Few studies have detected, or focused on methods to detect, HIV in wastewater (Ansari et al.,
402 1992; Kadadou et al., 2022; McCall et al., 2020; Preston et al., 1991). Wastewater surveillance of
403 HIV in high-risk areas could facilitate an increase in status awareness, early application of
404 retroactive and antiviral drugs, and decrease transmission of the viruses in communities. Indeed,
405 it is important to consider the contribution of adults (e.g., teachers, faculty, administrators) when
406 it comes to assessing the prevalence of viruses and age-related disease burdens in school samples,
407 nonetheless wastewater monitoring can be utilized to understand the circulation of clinically
408 important viral pathogens affecting K-12 populations.

409
410 Orthohepadnavirus, cytomegalovirus, and, mupapillomavirus were the only genera detected in
411 building-level samples that were undetected in WWTP samples. Cytomegalovirus and
412 mupapillomavirus are associated with retinal lesions in immunocompromised individuals
413 (Dioverti and Razonable, 2016) and skin warts (McBride, 2022) were only detected in nursing
414 home E and the jail samples, respectively. Albeit detected in 4 building-level samples, the jail was

415 the only sample that contained reads homologous to the hepatitis B virus (HBV) in the
416 orthohepadnavirus genus (data not shown). HBV is the causative agent of hepatitis B and can
417 remain in the body for prolonged periods of time. It is primarily transmitted through serum and
418 blood and is the most common chronic viral infection worldwide (Trépo et al., 2014). Incarcerated
419 populations have a disproportionate HBV burden, with an estimated seroprevalence of currently
420 or previously infected inmates of > 40% in the U.S. (Smith et al., 2017). HBV is rarely detected
421 in wastewater samples (Symonds et al., 2009), which suggests that sensitive, targeted
422 metagenomics can be used as a screening tool for both rare and abundant viruses for public health
423 surveillance.

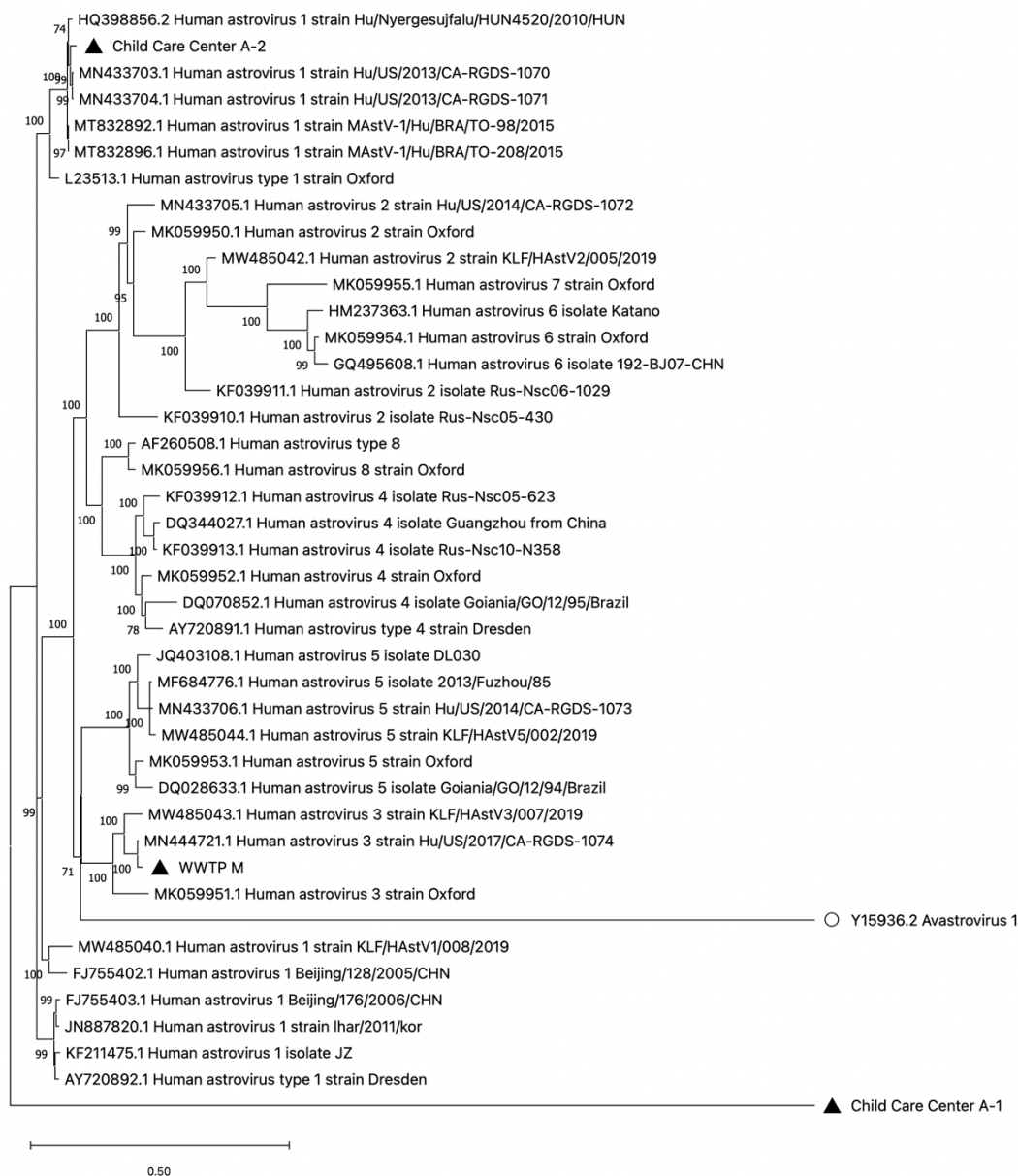
424

425 **3.4. Genetic diversity of astroviruses in wastewater**

426

427 Near complete and whole-genome contigs were clustered with HAstV-1 and -3 reference
428 sequences (Figure 5). HAstV-1 is one of the most common serotypes isolated from pediatric
429 patients (Cortez et al., 2017; de Grazia et al., 2004; Li et al., 2010) and is frequently detected in
430 wastewater (Hata et al., 2018; Meleg et al., 2006; Ng et al., 2012; Zhou et al., 2014). Although
431 more studies are needed to thoroughly understand differences in virus behaviors between HAstV
432 serotypes, one study found HAstV-3 to be associated with higher fecal concentrations and more
433 persistent and severe gastroenteritis as compared to the other HAstVs (Caballero et al., 2003).
434 Although detected less frequently than HAstV-1, HAstV-3 has also been detected in wastewater
435 samples (Hata et al., 2018). The Child Care Center A-1 contig did not form a cluster with any of
436 the reference sequences. This could be due to further assembly errors or the presence of a divergent

437 sequence. Nonetheless, targeted metagenomics was able to generate putative whole genome
438 contigs at the species level for astroviruses and identify serotypes in wastewater samples.



439
440

441 Figure 5. Maximum likelihood phylogenetic tree based on complete genomes of mamastrovirus
442 1 using the generalized time-reversible model in Fasttree 2.1.11. The tree was constructed with
443 38 reference sequences from GenBank and 3 putative near complete or complete whole genome
444 contigs from this study. Contigs are denoted with a black triangle. Avastrovirus 1 (white circle)

445 was used as an outgroup. The number on the branches represent the bootstrap percentage and the
446 scale bar refers to the phylogenetic distance in nucleotide substitutions per site. MEGA 11.0 was
447 used for tree visualization only.

448

449 **3.5. Limitations and implications**

450

451 While we were able to identify important human viruses in building-level samples, more samples
452 of the same community type are needed to statistically evaluate associations between community
453 type and viruses detected. For this reason, determining statistical associations between the viruses
454 detected and demographic groups was outside the scope of this study. However, virus profiles
455 noticeably varied among building-level samples and several viruses detected at the building-level,
456 including HBV, were not detected in WWTP samples. This suggests that building-level sampling
457 can increase the chances of detecting viruses not commonly detected in centralized wastewater
458 samples. Overall, targeted metagenomics increased the number of reads and unique vertebrate viral
459 families in wastewater samples when compared to untargeted metagenomics. Additionally,
460 species-level astrovirus reads produced putative whole-genome contigs with homologies with
461 HAstV-1 and -3 serotypes. This study demonstrates that targeted metagenomic sequencing of
462 WWTP and building-level samples can be used as an initial screening step for public health
463 surveillance and potentially to classify abundant viruses at the subspecies level.

464

465 **Acknowledgment**

466 We'd like to thank Houston Public Works and the Houston Health Department for providing
467 wastewater samples. We'd also like to thank Bryce Kille and Advait Balaji in Dr. Todd Treangen's

468 group for advising on the homology and phylogenetic analyses of astroviruses. Thanks to the
469 Genome Technology Access Center at McDonnell Genome Institute team at Washington
470 University, and specifically Elizabeth Appelbaum and Andrew Emory for providing targeted and
471 untargeted wastewater sequencing data. Lastly, we'd like to thank Madeline Wolken, Lauren
472 Bauhs, Russell Carlson, Kyle Palmer, and Whitney Rich in Lauren Stadler's lab group at Rice
473 University for their assistance with wastewater sample collection and processing.

474

475 **Funding Sources**

476 This work was supported by the Houston Health Department and grants from the National Science
477 Foundation (CBET 2029025) and seed funds from Rice University.

478

479 **Supporting Information**

480 Appendix 1: Protocol for Two-Strand cDNA Reverse Transcription for Targeted Metagenomic
481 Sequencing

482

483 Appendix 2: Supplementary Table and Figures

484

485 Figure S1. Integrative Genomics Viewer (IGV) plot of BWA-MEM read coverage alignment for
486 putative whole genome contig for the Child Care Center A-1 sample.

487

488 Figure S2. Integrative Genomics Viewer (IGV) plot of BWA-MEM read coverage alignment for
489 putative whole genome contig for the Child Care Center A-2 sample.

490

491 Figure S3. Integrative Genomics Viewer (IGV) plot of BWA-MEM read coverage alignment for
492 putative whole genome contig for the WWTP M sample.

493

494 Figure S4. Number of unique vertebrate families detected between electronegative membrane
495 filtered (MF), directly extracted (DE), targeted with probe-based capture, and untargeted
496 samples.

497

498 Figure S5. Percent of vertebrate reads detected between electronegative membrane filtered (MF),
499 directly extracted (DE), targeted with probe-based capture, and untargeted samples.

500

501 Table S1. Summary of contig coverage results from alignment with BWA-MEM and
502 visualization with Integrative Genomics Viewer (IGV).

503

504 Table S2. Summary of read count per vertebrate virus family in untargeted/electronegative
505 membrane filtered (MF) and targeted/MF samples for WWTPs M and O.

506

507 Table S3. Summary of read count per vertebrate virus family in targeted/directly extracted (DE)
508 and targeted/electronegative membrane filtered (MF) samples for WWTPs M and O.

509

510 Table S4. Total number of sequencing reads per sample assigned to viruses using the ViroMatch
511 pipeline.

512

513 Table S5. Summary human viruses detected in building-level and WWTP samples and their
514 associated route of transmission.

515

516 **References**

- 517 Ansari, S.A., Farrah, S.R., Chaudhry, G.R., 1992. Presence of Human Immunodeficiency Virus
518 Nucleic Acids in Wastewater and Their Detection by Polymerase Chain Reaction. *Appl*
519 *Environ Microbiol* 58, 3984–3990.
- 520 Aw, T.G., Howe, A., Rose, J.B., 2014. Metagenomic approaches for direct and cell culture
521 evaluation of the virological quality of wastewater. *J Virol Methods* 210, 15–21.
522 <https://doi.org/10.1016/j.jviromet.2014.09.017>
- 523 Barrios, R.E., Lim, C., Kelley, M.S., Li, X., 2021. SARS-CoV-2 concentrations in a wastewater
524 collection system indicated potential COVID-19 hotspots at the zip code level. *Science of*
525 *the Total Environment* 800. <https://doi.org/10.1016/j.scitotenv.2021.149480>
- 526 Bibby, K., Peccia, J., 2013. Identification of viral pathogen diversity in sewage sludge by
527 metagenome analysis. *Environ Sci Technol* 47, 1945–1951.
528 <https://doi.org/10.1021/es305181x>
- 529 Bisseux, M., Colombet, J., Mirand, A., Roque-Afonso, A.M., Abravanel, F., Izopet, J.,
530 Archimbaud, C., Peigue-Lafeuille, H., Debroas, D., Bailly, J.L., Henquell, C., 2018.
531 Monitoring human enteric viruses in wastewater and relevance to infections encountered in
532 the clinical setting: A one-year experiment in central France, 2014 to 2015.
533 *Eurosurveillance* 23, 1–11. <https://doi.org/10.2807/1560-7917.ES.2018.23.7.17-00237>
- 534 Caballero, S., Guix, S., El-Senousy, W.M., Calicó, I., Pintó, R.M., Bosch, A., 2003. Persistent
535 gastroenteritis in children infected with astrovirus: Association with serotype-3 strains. *J*
536 *Med Virol* 71, 245–250. <https://doi.org/10.1002/jmv.10476>
- 537 Cantalupo, P.G., Calgua, B., Zhao, G., Hundesa, A., Wier, A.D., Katz, J.P., Grabe, M., Hendrix,
538 R.W., Girones, R., Wang, D., Pipas, J.M., 2011. Raw sewage harbors diverse viral
539 populations. *mBio* 2, 1–11. <https://doi.org/10.1128/mBio.00180-11>
- 540 Choi, P.M., Tschärke, B., Samanipour, S., Hall, W.D., Gartner, C.E., Mueller, J.F., Thomas, K.
541 v., O'Brien, J.W., 2019. Social, demographic, and economic correlates of food and
542 chemical consumption measured by wastewater-based epidemiology. *Proc Natl Acad Sci U*
543 *S A* 116, 21864–21873. <https://doi.org/10.1073/pnas.1910242116>
- 544 Cortez, V., Freiden, P., Gu, Z., Adderson, E., Hayden, R., Schultz-Cherry, S., 2017. Persistent
545 infections with diverse co-circulating astroviruses in pediatric oncology patients, Memphis,
546 Tennessee, USA. *Emerg Infect Dis* 23, 288–290. <https://doi.org/10.3201/eid2302.161436>
- 547 de Grazia, S., Giammanco, G.M., Colomba, C., Cascio, A., Arista, S., 2004. Molecular
548 epidemiology of astrovirus infection in Italian children with gastroenteritis. *Clinical*
549 *Microbiology and Infection* 10, 1025–1029. <https://doi.org/10.1111/j.1469-0691.2004.00995.x>
- 550
- 551 Deeks, S.G., Overbaugh, J., Phillips, A., Buchbinder, S., 2015. HIV infection. *Nat Rev Dis*
552 *Primers* 1. <https://doi.org/10.1038/nrdp.2015.35>
- 553 Dioverti, M.V., Razonable, R.R., 2016. Cytomegalovirus. *Microbiol Spectr* 4.
554 <https://doi.org/10.1128/microbiolspec.DMIH2-0022-2015>

- 555 Donato, C., Vijaykrishna, D., 2017. The broad host range and genetic diversity of mammalian
556 and avian astroviruses. *Viruses*. <https://doi.org/10.3390/v9050102>
- 557 Farkas, K., Cooper, D.M., McDonald, J.E., Malham, S.K., de Rougemont, A., Jones, D.L., 2018.
558 Seasonal and spatial dynamics of enteric viruses in wastewater and in riverine and estuarine
559 receiving waters. *Science of the Total Environment* 634, 1174–1183.
560 <https://doi.org/10.1016/j.scitotenv.2018.04.038>
- 561 Fernandez-Cassi, X., Martínez-Puchol, S., Silva-Sales, M., Cornejo, T., Bartolome, R., Bofill-
562 Mas, S., Girones, R., 2020. Unveiling viruses associated with gastroenteritis using a
563 metagenomics approach. *Viruses* 12. <https://doi.org/10.3390/v12121432>
- 564 Fernandez-Cassi, X., Scheidegger, A., Bänziger, C., Cariti, F., Tuñas Corzon, A.,
565 Ganesanandamoorthy, P., Lemaitre, J.C., Ort, C., Julian, T.R., Kohn, T., 2021. Wastewater
566 monitoring outperforms case numbers as a tool to track COVID-19 incidence dynamics
567 when test positivity rates are high. *Water Res* 200.
568 <https://doi.org/10.1016/j.watres.2021.117252>
- 569 Galani, A., Aalizadeh, R., Kostakis, M., Markou, A., Alygizakis, N., Lytras, T., Adamopoulos,
570 P.G., Peccia, J., Thompson, D.C., Kontou, A., Karagiannidis, A., Lianidou, E.S., Avgeris,
571 M., Paraskevis, D., Tsiodras, S., Scorilas, A., Vasiliou, V., Dimopoulos, M.A., Thomaidis,
572 N.S., 2022. SARS-CoV-2 wastewater surveillance data can predict hospitalizations and ICU
573 admissions. *Science of the Total Environment* 804.
574 <https://doi.org/10.1016/j.scitotenv.2021.150151>
- 575 Girón-Guzmán, I., Díaz-Reolid, A., Cuevas-Ferrando, E., Falcó, I., Cano-Jiménez, P., Comas, I.,
576 Pérez-Cataluña, A., Sánchez, G., 2023. Evaluation of two different concentration methods
577 for surveillance of human viruses in sewage and their effects on SARS-CoV-2 sequencing.
578 *Science of the Total Environment* 862. <https://doi.org/10.1016/j.scitotenv.2022.160914>
- 579 Hall, R.J., Wang, J., Todd, A.K., Bissielo, A.B., Yen, S., Strydom, H., Moore, N.E., Ren, X.,
580 Huang, Q.S., Carter, P.E., Peacey, M., 2014. Evaluation of rapid and simple techniques for
581 the enrichment of viruses prior to metagenomic virus discovery. *J Virol Methods* 195, 194–
582 204. <https://doi.org/10.1016/j.jviromet.2013.08.035>
- 583 Haramoto, E., Kitajima, M., Hata, A., Torrey, J.R., Masago, Y., Sano, D., Katayama, H., 2018. A
584 review on recent progress in the detection methods and prevalence of human enteric viruses
585 in water. *Water Res* 135, 168–186. <https://doi.org/10.1016/j.watres.2018.02.004>
- 586 Hata, A., Kitajima, M., Haramoto, E., Lee, S., Ihara, M., Gerba, C.P., Tanaka, H., 2018. Next-
587 generation amplicon sequencing identifies genetically diverse human astroviruses, including
588 recombinant strains, in environmental waters. *Sci Rep* 8. <https://doi.org/10.1038/s41598-018-30217-y>
- 589 Hellmér, M., Paxéus, N., Magnus, L., Enache, L., Arnholm, B., Johansson, A., Bergström, T.,
590 Norder, H., 2014. Detection of pathogenic viruses in sewage provided early warnings of
591 hepatitis A virus and norovirus outbreaks. *Appl Environ Microbiol* 80, 6771–6781.
592 <https://doi.org/10.1128/AEM.01981-14>
- 593 Hjelmsø, M.H., Hellmér, M., Fernandez-Cassi, X., Timoneda, N., Lukjancenka, O., Seidel, M.,
594 Elsässer, D., Aarestrup, F.M., Löfström, C., Bofill-Mas, S., Abril, J.F., Girones, R., Schultz,
595 A.C., 2017. Evaluation of methods for the concentration and extraction of viruses from
596 sewage in the context of metagenomic sequencing. *PLoS One* 12, 1–17.
597 <https://doi.org/10.1371/journal.pone.0170199>
- 598 Hopkins, L., Persse, D., Caton, K., Ensor, K., Schneider, R., McCall, C., Stadler, L.B., 2023.
599 Citywide wastewater SARS-CoV-2 levels strongly correlated with multiple disease
600

- 601 surveillance indicators and outcomes over three COVID-19 waves. *Science of the Total*
602 *Environment* 855. <https://doi.org/10.1016/j.scitotenv.2022.158967>
- 603 Kadadou, D., Tizani, L., Wadi, V.S., Banat, F., Alsafar, H., Yousef, A.F., Barceló, D., Hasan,
604 S.W., 2022. Recent advances in the biosensors application for the detection of bacteria and
605 viruses in wastewater. *J Environ Chem Eng* 10. <https://doi.org/10.1016/j.jece.2021.107070>
- 606 Kirby, A.E., Walters, M.S., Jennings, W.C., Fugitt, R., LaCross, N., Mattioli, M., Marsh, Z.A.,
607 Roberts, V.A., Mercante, J.W., Yoder, J., Hill, V.R., 2021. Using Wastewater Surveillance
608 Data to Support the COVID-19 Response — United States, 2020–2021. *Morbidity and*
609 *Mortality Weekly Report* 70, 1242–1244.
- 610 LaTurner, Z.W., Zong, D.M., Kalvapalle, P., Gamas, K.R., Terwilliger, A., Crosby, T., Ali, P.,
611 Avadhanula, V., Santos, H.H., Weesner, K., Hopkins, L., Piedra, P.A., Maresso, A.W.,
612 Stadler, L.B., 2021. Evaluating recovery, cost, and throughput of different concentration
613 methods for SARS-CoV-2 wastewater-based epidemiology. *Water Res* 197.
614 <https://doi.org/10.1016/j.watres.2021.117043>
- 615 Li, C.Y., Liu, N., Guo, W.D., Yu, Q., Wang, W.R., Song, Z.Z., Yan, H., Luo, Y., Lu, A.T., Li,
616 H.Y., Zhu, L., Duan, Z.J., 2010. Outbreak of neonatal gastroenteritis associated with
617 astrovirus serotype 1 at a hospital in inner Mongolia, China. *J Clin Microbiol* 48, 4306–
618 4309. <https://doi.org/10.1128/JCM.00797-10>
- 619 Lyman, W.H., Walsh, J.F., Kotch, J.B., Weber, D.J., Gunn, E., Vinjé, J., 2009. Prospective Study
620 of Etiologic Agents of Acute Gastroenteritis Outbreaks in Child Care Centers. *Journal of*
621 *Pediatrics* 154, 253–257. <https://doi.org/10.1016/j.jpeds.2008.07.057>
- 622 Martínez-Puchol, S., Rusiñol, M., Fernández-Cassi, X., Timoneda, N., Itarte, M., Andrés, C.,
623 Antón, A., Abril, J.F., Girones, R., Bofill-Mas, S., 2020. Characterisation of the sewage
624 virome: comparison of NGS tools and occurrence of significant pathogens. *Science of the*
625 *Total Environment* 713. <https://doi.org/10.1016/j.scitotenv.2020.136604>
- 626 Matus, M., Duvallat, C., Soule, M.K., Kearney, S.M., Endo, N., Ghaeli, N., Brito, I., Ratti, C.,
627 Kujawinski, E.B., Alm, E.J., 2019. 24-hour multi-omics analysis of residential sewage
628 reflects human activity and informs public health. *bioRxiv*. <https://doi.org/10.1101/728022>
- 629 McBride, A.A., 2022. Human papillomaviruses: diversity, infection and host interactions. *Nat*
630 *Rev Microbiol* 20, 95–108. <https://doi.org/10.1038/s41579-021-00617-5>
- 631 McCall, C., Wu, H., Miyani, B., Xagorarakis, I., 2020. Identification of multiple potential viral
632 diseases in a large urban center using wastewater surveillance. *Water Res* 184.
633 <https://doi.org/10.1016/j.watres.2020.116160>
- 634 McCall, C., Wu, H., O'Brien, E., Xagorarakis, I., 2021. Assessment of enteric viruses during a
635 hepatitis outbreak in Detroit MI using wastewater surveillance and metagenomic analysis. *J*
636 *Appl Microbiol* 131, 1539–1554. <https://doi.org/10.1111/jam.15027>
- 637 Meleg, E., Jakab, F., Kocsis, B., Bányai, K., Melegh, B., Szucs, G., 2006. Human astroviruses in
638 raw sewage samples in Hungary. *J Appl Microbiol*. <https://doi.org/10.1111/j.1365-2672.2006.02997.x>
- 639
- 640 Mondal, S., Feirer, N., Brockman, M., Preston, M.A., Teter, S.J., Ma, D., Goueli, S.A., Moorji,
641 S., Saul, B., Cali, J.J., 2021. A direct capture method for purification and detection of viral
642 nucleic acid enables epidemiological surveillance of SARS-CoV-2. *Science of the Total*
643 *Environment* 795. <https://doi.org/10.1016/j.scitotenv.2021.148834>
- 644 Montoya, A., Mody, L., 2011. Common infections in nursing homes: A review of current issues
645 and challenges. *Aging health* 7, 889–899. <https://doi.org/10.2217/ahe.11.80>

- 646 Newton, R.J., McLellan, S.L., Dila, D.K., Vineis, J.H., Morrison, H.G., Murat Eren, A., Sogin,
647 M.L., 2015. Sewage reflects the microbiomes of human populations. *mBio* 6.
648 <https://doi.org/10.1128/mBio.02574-14>
- 649 Ng, T.F.F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., Oderinde,
650 B.S., Wommack, K.E., Delwart, E., 2012. High variety of known and new RNA and DNA
651 viruses of diverse origins in untreated sewage. *J Virol* 86, 12161–12175.
652 <https://doi.org/10.1128/jvi.00869-12>
- 653 O'Brien, E., Nakyazze, J., Wu, H., Kiwanuka, N., Cunningham, W., Kaneene, J.B., Xagorarakis,
654 I., 2017. Viral diversity and abundance in polluted waters in Kampala, Uganda. *Water Res*
655 127, 41–49. <https://doi.org/10.1016/j.watres.2017.09.063>
- 656 Paskey, A.C., Frey, K.G., Schroth, G., Gross, S., Hamilton, T., Bishop-Lilly, K.A., 2019.
657 Enrichment post-library preparation enhances the sensitivity of high-throughput
658 sequencing-based detection and characterization of viruses from complex samples. *BMC*
659 *Genomics* 20. <https://doi.org/10.1186/s12864-019-5543-2>
- 660 Peccia, J., Zulli, A., Brackney, D.E., Grubaugh, N.D., Kaplan, E.H., Casanovas-Massana, A.,
661 Ko, A.I., Malik, A.A., Wang, D., Wang, M., Warren, J.L., Weinberger, D.M., Arnold, W.,
662 Omer, S.B., 2020. Measurement of SARS-CoV-2 RNA in wastewater tracks community
663 infection dynamics. *Nat Biotechnol* 38, 1164–1167. [https://doi.org/10.1038/s41587-020-](https://doi.org/10.1038/s41587-020-0684-z)
664 [0684-z](https://doi.org/10.1038/s41587-020-0684-z)
- 665 Preston, D.R., Farrah, S.R., Bitton, G., Chaudhry, G.Rasul., 1991. Detection of nucleic acids
666 homologous to human immunodeficiency virus in wastewater". *J Virol Methods* 33, 383–
667 390.
- 668 Prevost, B., Lucas, F.S., Goncalves, A., Richard, F., Moulin, L., Wurtzer, S., 2015. Large scale
669 survey of enteric viruses in river and waste water underlines the health status of the local
670 population. *Environ Int* 79, 42–50. <https://doi.org/10.1016/j.envint.2015.03.004>
- 671 R Core Team, 2022. R: A language and environment for statistical computing. Vienna, Austria.
- 672 Segelhurst, E., Bard, J.E., Pillsbury, A.N., Mahbulul Alam, M., Lamb, A., Zhu, C., Pohlman, A.,
673 Boccolucci, A., Emerson, J., Yergeau, D.A., Nowak, N.J., Bradley, I.M., Surtees, J.A., Ye,
674 Y., 2022. Improved Robustness of SARS-CoV-2 Whole-Genome Sequencing from
675 Wastewater with a Nonselective Virus Concentration Method. *medRxiv*.
676 <https://doi.org/10.1101/2022.09.07.22279692>
- 677 Smith, J.M., Uvin, A.Z., Macmadu, A., Rich, J.D., 2017. Epidemiology and Treatment of
678 Hepatitis B in Prisoners. *Curr Hepatol Rep* 16, 178–183. [https://doi.org/10.1007/s11901-](https://doi.org/10.1007/s11901-017-0364-8)
679 [017-0364-8](https://doi.org/10.1007/s11901-017-0364-8)
- 680 Spurbeck, R.R., Minard-Smith, A., Catlin, L., 2021. Feasibility of neighborhood and building
681 scale wastewater-based genomic epidemiology for pathogen surveillance. *Science of the*
682 *Total Environment* 789. <https://doi.org/10.1016/j.scitotenv.2021.147829>
- 683 Sutton, T.D.S., Clooney, A.G., Ryan, F.J., Ross, R.P., Hill, C., 2019. Choice of assembly
684 software has a critical impact on virome characterisation. *Microbiome* 7.
685 <https://doi.org/10.1186/s40168-019-0626-5>
- 686 Symonds, E.M., Griffin, D.W., Breitbart, M., 2009. Eukaryotic viruses in wastewater samples
687 from the United States. *Appl Environ Microbiol* 75, 1402–1409.
688 <https://doi.org/10.1128/AEM.01899-08>
- 689 Tao, Z., Lin, X., Liu, Y., Ji, F., Wang, S., Xiong, P., Zhang, L., Xu, Q., Xu, A., Cui, N., 2022.
690 Detection of multiple human astroviruses in sewage by next generation sequencing. *Water*
691 *Res* 218. <https://doi.org/10.1016/j.watres.2022.118523>

- 692 Trépo, C., Chan, H.L.Y., Lok, A., 2014. Hepatitis B virus infection. *The Lancet* 384, 2053–2063.
693 [https://doi.org/10.1016/S0140-6736\(14\)60220-8](https://doi.org/10.1016/S0140-6736(14)60220-8)
- 694 Vu, D.L., Bosch, A., Pintó, R.M., Guix, S., 2017. Epidemiology of classic and novel human
695 astrovirus: Gastroenteritis and beyond. *Viruses* 9. <https://doi.org/10.3390/v9020033>
- 696 Wang, D., Urisman, A., Liu, Y.T., Springer, M., Ksiazek, T.G., Erdman, D.D., Mardis, E.R.,
697 Hickenbotham, M., Magrini, V., Eldred, J., Latreille, J.P., Wilson, R.K., Ganem, D.,
698 DeRisi, J.L., 2003. Viral discovery and sequence recovery using DNA microarrays. *PLoS*
699 *Biol* 1. <https://doi.org/10.1371/journal.pbio.0000002>
- 700 Westerhoff, P., Lee, S., Yang, Y., Gordon, G.W., Hristovski, K., Halden, R.U., Herckes, P.,
701 2015. Characterization, Recovery Opportunities, and Valuation of Metals in Municipal
702 Sludges from U.S. Wastewater Treatment Plants Nationwide. *Environ Sci Technol* 49,
703 9479–9488. <https://doi.org/10.1021/es505329q>
- 704 Wurtzer, S., Waldman, P., Levert, M., Cluzel, N., Almayrac, J.L., Charpentier, C., Masnada, S.,
705 Gillon-Ritz, M., Mouchel, J.M., Maday, Y., Boni, M., Moulin, L., Marechal, V., le
706 Guyader, S., Bertrand, I., Gantzer, C., Descamps, D., Charpentier, Charlotte, Houhou, N.,
707 Visseaux, B., Calvez, V., Marcelin, A.G., Marot, S., Jary, A., Chaix, M.L., Delaugerre, C.,
708 Salmona, M., Legoff, J., Gault, E., Rameix-Welti, M.A., L'honneur, A.S., Mariaggi, A.A.,
709 Rozenberg, F., Roque, A.M., Mouna, L., Fenaux, H., Leruez-Ville, M., Morand-Joubert, L.,
710 Lambert-Niclot, S., Schnuriger, A., Dubois, C., Bourgeois-Nicolaos, N., Brichler, S.,
711 Delagrèverie, H., Bouvier-Alias, M., Fourati, S., Rodriguez, C., Pawlotsky, J.M., 2022.
712 SARS-CoV-2 genome quantification in wastewaters at regional and city scale allows
713 precise monitoring of the whole outbreaks dynamics and variants spreading in the
714 population. *Science of the Total Environment* 810.
715 <https://doi.org/10.1016/j.scitotenv.2021.152213>
- 716 Wylie, K.M., Wylie, T.N., Buller, R., Herter, B., Cannella, M.T., Storcha, G.A., 2018. Detection
717 of Viruses in Clinical Samples by Use of Metagenomic Sequencing and Targeted Sequence
718 Capture. *J Clin Microbiol* 56. <https://doi.org/10.1128/JCM.01123-18>
- 719 Wylie, T.N., Wylie, K.M., 2021. ViroMatch: A Computational Pipeline for the Detection of
720 Viral Sequences from Complex Metagenomic Data. *Microbiol Resour Announc* 10.
721 <https://doi.org/10.1128/mra.01468-20>
- 722 Wylie, T.N., Wylie, K.M., Herter, B.N., Storch, G.A., 2015. Enhanced virome sequencing using
723 targeted sequence capture. *Genome Res* 25, 1910–1920.
724 <https://doi.org/10.1101/gr.191049.115>
- 725 Xu, B., Liu, L., Huang, X., Ma, H., Zhang, Y., Du, Y., Wang, P., Tang, X., Wang, H., Kang, K.,
726 Zhang, S., Zhao, G., Wu, W., Yang, Y., Chen, H., Mu, F., Chen, W., 2011. Metagenomic
727 analysis of fever, thrombocytopenia and leukopenia syndrome (FTLS) in Henan province,
728 China: Discovery of a new bunyavirus. *PLoS Pathog* 7.
729 <https://doi.org/10.1371/journal.ppat.1002369>
- 730 Zhao, L., Zou, Y., Li, Y., Miyani, B., Spooner, M., Gentry, Z., Jacobi, S., David, R.E.,
731 Withington, S., McFarlane, S., Faust, R., Sheets, J., Kaye, A., Broz, J., Gosine, A., Mobley,
732 P., Busch, A.W.U., Norton, J., Xagorarakis, I., 2022. Five-week warning of COVID-19
733 peaks prior to the Omicron surge in Detroit, Michigan using wastewater surveillance.
734 *Science of the Total Environment* 844. <https://doi.org/10.1016/j.scitotenv.2022.157040>
- 735 Zhou, N., Lin, X., Wang, S., Wang, H., Li, W., Tao, Z., Xu, A., 2014. Environmental
736 surveillance for human astrovirus in Shandong Province, China in 2013. *Sci Rep* 4.
737 <https://doi.org/10.1038/srep07539>

738

739 **For Table of Contents Only**

740

