bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 *Article*

2 A New High-throughput Screening (HiTS) Method 3 for Phages – Enabling Crude Isolation and Fast 4 Identification of Diverse Phages with Therapeutic 5 Potential

6 **Nikoline S. Olsen [1], Niels Bohse Hendriksen[1], Lars H. Hansen [2,‡,*] and Witold Kot [2,‡,*]**

7 [1] Department of Environmental Science, Aarhus University, Frederiksborgvej 399, Roskilde, Denmark;
8 niso@envs.au.dk
9 [2] Department of Plant and Environmental Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871
10 Frederiksberg C, Denmark; lhha@plen.ku.dk, wk@plen.ku.dk
11 ‡ W. K. and L. H. H. initiated and supervised the project equally.
12 * Correspondence: lhha@plen.ku.dk **Phone:** +45 28 75 20 53, wk@plen.ku.dk **Phone:** +45 35 33 38 77
13 Received: date; Accepted: date; Published: date

14 **Abstract:** Bacteriophage therapy and application of phages as biocontrol in plant production and
15 food processing, all necessitates acquisition of suitable phages. Depending on purpose, the selection
16 criteria of phage characteristics include lifestyle (lytic/lysogenic), host range, physical stability and
17 absence of unwanted genetic traits such as integrases, antibiotic resistance or bacterial virulence
18 factors. The exclusivity of antibiotic resistant clinical infections and possible development of phage-
19 resistance instigates a need to continually build sizeable phage libraries and also be able to rapidly
20 isolate and characterise novel phages of specified bacterial hosts. Current methods for phage
21 isolation are both laborious and time consuming, suitable only for the isolation of a limited number
22 of phages. Thus, we developed the ***Hi****gh-****T****hroughput* ***S****creening* (HITS) method for phages for fast
23 isolation and identification of potentially hundreds of distinct phages against single hosts. This
24 scalable method enables screening of hundreds of samples, in multiple simultaneous setups with
25 varying parameters increasing the likelihood of isolating multiple distinct phages specific for the
26 given conditions. The efficiency of the method is emphasised by our screening of 200 environmental
27 samples, resulting in the identification of an abundance of unique phage species lytic to *Escherichia*
28 *coli*, *Salmonella Enterica, Enterococcus faecalis* and *Pseudomonas aeruginosa*.

29 **Keywords:** bacteriophage; phage isolation; phagetherapy; high-throughput; wastewater; *Escherichia*
30 *coli*, *Salmonella Enterica, Enterococcus faecalis*; *Pseudomonas aeruginosa*.
31

---

32 1. Introduction

33 The upsurge of antibiotic resistant bacteria is one of the main health concerns of our time [1].
34 Pathogenic bacterial infections are becoming ever more difficult to treat, and even last resort
35 antibiotics such as the glycopeptide antibiotics vancomycin and teicoplanin are falling short as
36 efficient antimicrobial agents [2]. Bacteria are consecutively acquiring antibiotic resistance and
37 develop multidrug resistance [3], which necessitates the development of alternative antimicrobials
38 or means to increase the efficiency of existing antibiotics. Phage therapy (PT) is the therapeutic use
39 of the viral antagonists of bacteria, the bacteriophages (phages), to treat bacterial infections in humans
40 or animals. Most bacteriophages have narrow host-ranges, limiting their infectivity to specific species
41 or even strains. Consequently, PT does not instigate drastic perturbations of natural microbiota like
42 traditional antibiotic treatments [4]. Though studies have been limited, PT has not been shown to
43 have any adverse side effects [5]. Moreover, PT has shown potential as a last resort treatment of multi-

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

44 resistant bacterial infections, when traditional antibiotics fall short [6]–[8]. Hence, PT is, especially
45 when applied as a combination therapy together with conventional antimicrobials, foreseen to play
46 an essential role in the multifaceted strategy required to combat the lurking antibiotic crisis [1], [9].
47 Furthermore, the use of phages for biocontrol in plant production and food processing has displayed
48 a promising potential [10], and could be a sustainable alternative to traditional chemicals facing
49 restrictions due to concerns for public health and the spread of resistance [11].
50 Yet, a successful biocontrol or PT venture requires phages with different modes of action, and
51 lots of them. Infection-specific phages and prepared phage cocktails are rarely generalisable [12].
52 Clinical infections can be unparalleled and call for *de novo* isolation or genetic engineering, as was the
53 recent case with a 15-year old patient with cystic fibrosis caused by *Mycobacterium abscessus* [6]. More
54 than 10 000 phages infecting *Mycobacterium smegmatis* were screened in addition to 100 environmental
55 samples, resulting in only three suitable phages, two of them requiring genetic engineering [6].
56 Indeed, one of the greater hurdles for effective PT, is the availability of suitable phages [12].
57 Methodologies for isolation of phages have not changed much since phages were discovered more
58 than 100 years ago. The procedures are laborious and time-consuming. In general, phages are isolated
59 by either direct plating or by enrichment and then subsequent purification. Enrichment entails an
60 introduction of a host to a phage-containing sample, which is afterwards removed by centrifugation
61 and filtration when phages have been amplified. Direct plating and purification is typically
62 performed with the soft-agar overlay technique, first described by A. Gratia in 1936 [13].
63 Improvements to increase throughputs have been proposed, such as tube-free agar overlays [14], and
64 phage activity can now be measured by means more suitable for automation, like colorimetric
65 methods [15]. However, no truly high-throughput isolation method has, to our knowledge been
66 offered. A citizen science approach, like the great effort performed by The Science Education Alliance
67 Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) has resulted in the
68 isolation of thousands of phages against *M. smegmatis* [16]. But this type of approach requires both
69 substantial funding and facilities.
70 In order to establish and expand libraries of phages relevant for PT and biocontrol, affordable,
71 fast and efficient screening methods are needed to enable rapid isolation and identification of
72 candidate phages. Large libraries of phages infecting the same single host, also enables important
73 phage-host interaction studies, expanding our understanding of phage taxonomy and ecology. Here
74 we present the ***H**igh-**T**hroughput **S**creening* (HiTS) method for phages, which enables a single person
75 to go from a high number of samples to a plethora of identified phages within weeks. The simplicity
76 of the method enables >500 samples to be handled simultaneously. The HiTS method selects for
77 predominantly lytic and easily culturable phages. The resolution is a single or a few phages from
78 each sample processed. The integrated sequencing of the identified phages allows for an early
79 assessment of genomes enabling the selection of candidates which do not possess any unwanted
80 genetic-traits and are thus suitable for further characterisation and potential application as PT or
81 biocontrol agents.

82 ## 2. Materials and Methods

83 The method presented is host-system independent and can thus be applied for screening of
84 environmental samples for phages virulent to any culturable fast-growing aerobic or facultative
85 anaerobic bacteria by adjusting host media, incubation temperature and time. The protocol enables a
86 simple and fast (4 consecutive days), yet crude purification of single or a low number of distinctive
87 phages from a small sample volume (0.5-1.5 ml). The method allows for a high number of samples to
88 be handled, with simultaneous investigation of diverse sample matrices or parallel screenings of the
89 same sample-set with varying parameters e.g. host, pH, media, amendments and incubation
90 conditions (Figure 1). This increases the likelihood of sequestering multiple distinctive phages from
91 each sample. The method is suitable for both direct plaque sequencing (DPS) [17] and standard phage
92 DNA extraction from lysate. The screening procedure entails four steps: 1. *Phage amplification*, 2.
93 *Liquid purification*, 3. *Spot-test* and 4. *Phage collection and DPS or optional: plating of dilution series*.
94

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

95 *2.1. Protocol*

96 **High-throughput Screening (HiTS) Method for Phages**

97 The method, which is scalable for robotics, requires a centrifuge suitable for 96-well plates and
98 a 96-pin replicator or a one µl multichannel pipette. Multichannel pipettes or a pipetting robot may
99 ease many of the steps involved. Dry samples should be suspended prior to processing and
100 preferably centrifuged and filtered. The maximum number of samples per plate is 94. The sample
101 volume can be adjusted as desired and as applicable to available well plates. By increasing the volume
102 of raw sample input, the number of phages per incubation is also increased and thus the chance of
103 isolating these. Initial sample volume only affects step *1. Phage amplification*. Volumes and
104 concentrations suggested in step 0. and step 1. are suitable for screening 1.5 ml raw sample in deep-
105 well plates with a working volume (wV) of 2.2 ml (e.g. 732-0612, VWR, Radnor PA US). All
106 incubations should be performed under optimal host conditions (media and temperature) and hence
107 adjusted as required. In step 4 there is the option to either collect the phages and sequence them by
108 DPS or to do titers and aim for single plaques, and then do DPS or sequence phage-amplifications.
109

110 **0. Preparation**
111 a. Prepare all media, solutions and agar-plates.
112 b. Inoculate host cells in 2 × 10 ml liquid media and incubate overnight
113 c. (ON).
114

115 **1. Phage amplification (day 1)**
116 a. Distribute a maximum of 94 samples, in suitable volumes (1.5 ml) in
117 a deep-well plate (#1) with pierceable sealing tape (e.g. Z722529-
118 50EA, Excel Scientific, Victorville CA US). Sterilised water (1.5 ml) is
119 added as negative amplification-controls to wells D6 and E6. To each
120 of the 96 wells add:
121 90 µl $CaCl_2$ (0.25 M) and $MgCl_2$. (0.25 M), final conc. 10 mM
122 110 µl ON host culture, final conc. 5% V/V
123 500 µl Host media (conc. × 4.4), final conc. × 1
124 During addition of media, carefully pipette up and down a few times
125 to mix. Close the well-plate and incubate ON on a shaker (200 rpm) .
126 b. Inoculate ON host culture in 10 ml liquid media for next day.
127

128 **2. Liquid purification (day 2)**
129 a. Filter to remove host bacteria by transferring 200 µl (punch through
130 pierceable tape) from each well to a 96-well filter Plate (0.45 µm) (e.g.
131 MSHAS4510, Merck Millipore, Burlington MA US), pipette up and
132 down a few times before extracting. Centrifuge filter plate on top of
133 a new well-plate (#2, wV 200 ul e.g. 269787, Nunc, Roskilde, DK) at
134 900 x *g* for 2 minutes. Then add pierceable sealing tape to well-plate
135 #2. Discard the filter plate.
136 b. Prepare a third well-plate (#3, wV 200 µl) with pierceable sealing
137 tape and add:
138 180 µl host media (conc. x 1)
139 10 µl of host culture, final conc. 5% V/V
140 10 µl 0.2 M $CaCl_2$ and 0.2 M $MgCl_2$, final conc. 10 mM
141 Use the 96-pin replicator to transfer ~1 µl of each lysate (punch
142 through pierceable tape) in well-plate #2 to each well in well-plate
143 #3. Close well-plate #3 and incubate ON on a shaker (200 rpm). If
144 processing more than one set of samples, clean the 96-pin replicator
145 by ethanol and flame three times in between and make sure to cool
146 it down before re-use.

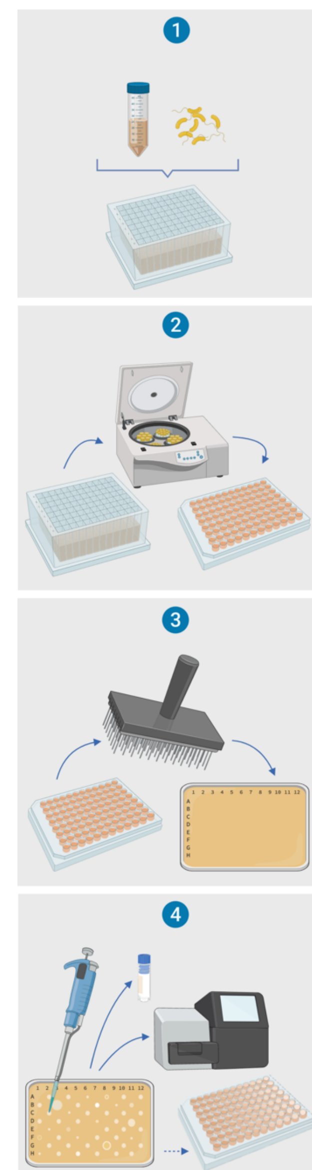


Figure 1 The four steps in the HiTS method. Illustration created with BioRender.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

147 **c.** Inoculate ON host culture in 10 ml liquid media for next day.
148
149 **3. Spot-test (day 3)**
150 a. Filter to remove host bacteria as described in step 2a.
151 b. Prepare two large (Ø 14 cm e.g. 82.1184.500, Sarstedt, Nürnbrecht DE or at least 12 x 8 cm e.g.
152 242811, Nunc) soft-agar overlay plates (A and B) of 0.5% agarose amended with:
153 $CaCl_2$ and $MgCl_2$ (final conc. 10 mM)
154 Host culture (final conc. 2.5 - 5 %)
155 c. While the plates solidify remove every second row of pipette-tips in a box of 200 µl pipette-tips
156 to facilitate the transfer of lysate from every second well in a chequered pattern into two new
157 microtiter plates with pierceable sealing tape, number #4 (A) and #5 (B) (Figure S1).
158 d. Use the 96-pin replicator to carefully transfer ~1 µl of lysates from well-plate A (#5) to the soft-
159 agar overlay plate A. Make sure to specify direction on the plate. The chequered pattern ensures
160 a safe distance between spotted samples, a negative amplification-control (D6 or E6) on each
161 plate and sterilisation-controls (every second tip) (Figure 1). Clean the 96-pin replicator by
162 ethanol and flame and repeat the procedure with well-plate B (#5) and soft-agar overlay plate B.
163 Incubate soft-agar overlay plates upside down ON. Seal well-plates A (#4) and B (#5) and store
164 at 4°C.
165 e. **Optional**: Inoculate ON host culture in 10 ml liquid media for titre estimation next day.
166
167 **4. Phage collection and DPS, or optional: plate dilution series (day 4)**
168 a. The centre of clearing zones (agar-plate A and B) is collected for DPS. Additional clearing zone
169 is dissolved in 100 µl SM-buffer [18], filtered (0.22 - 0.45 µm) and stored for future purification
170 and characterization. If clearing zones are too small for double collection make amplifications of
171 the phage-SM solutions (inoculate host bacteria in 10 ml media, after ~1 h add lysate, next day
172 centrifuge and filtrate) and extract DNA for sequencing from these.
173 b. Optional: titre the lysates by transferring phage solutions from positive (plaque-forming) wells
174 to new wells in column 1 of as many new well-plates (#6+, wV 200 µl) as required. Eight-fold
175 dilutions series are made within the well-plates by adding 180 µl SM-buffer to all wells in column
176 2-9 and then transferring 20 µl of the solution in column 1 to column 2 pipetting up and down
177 to mix and repeating the procedure for the remaining columns. Spot (~1 µl) the dilution series
178 on soft-agar overlay plates with a 96-pin replicator or multichannel pipette and incubate ON.
179 Next day: Count plaques or clearing zones for approximate titre. Do DPS of single plaques if
180 present and also collect clearing zone for phage storage. If single plaques are not present plate
181 lysate dilution giving rise to 10-50 plaques on a full plate by the soft-agar overlay method (add
182 the lysate to 4 ml 0.5% agarose with 10 mM $CaCl_2$ and $MgCl_2$ and 2.5-5% (V/V) ON host culture,
183 pour on standard petri dish with agar). Next day, pick diverting plaque morphologies for DPS
184 or phage-amplification, lysate hereof can be used for DNA extraction and phage storage.

185 *2.2. Phage screenings*

186 Five screenings were performed as described in *2.1 Protocol*, using *Escherichia coli, Salmonella*
187 *enterica, Enterococcus faecalis* or *Pseudomonas aeruginosa* as hosts (Table 1). For the *E. coli, E. faecalis* and
188 *S. enterica* screenings 188 distinct wastewater samples divided into two sets of 94 samples were used,
189 for the *P. aeruginosa* screening 82 wastewater samples were used together with eight soil samples and
190 four organic waste samples (Table S1). The *S. enterica* screenings were performed with both a small
191 sample volume (SV) of 0.5 ml and a large sample volume (LV) of 1.5 ml, while the *E. coli* and *E. faecalis*
192 screenings were only performed with 0.5 ml (SV) and the *P. aeruginosa* screening only with 1.5 ml
193 (LV). The SV screenings (*E. coli, E. faecalis* and *S. enterica)* followed the protocol, with the exception
194 that instead of DPS, lysates from step 3. corresponding to positive wells (those instigating clearing
195 zones) were used for DNA extraction and sequencing, while phages were stored by collecting top-
196 agar of clearing zones or plaques. In step 4 of the LV screenings (*S. enterica* and *P. aeruginosa*) lysates
197 from positive wells were tittered in the 96-well format and the most diluted lysates instigating single

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

198 plaques or clearing zones were used for making 10 ml amplification lysates for DNA extraction,
199 library preparation, sequencing and phage storage**.** All incubations were performed at 37°C.

200 2.2.1. Bacteria and growth media

201 The host bacteria used for phage screenings were *E. coli* (K-12, MG1655), *S. enterica subsp. enterica*
202 serovar Enteritidis PT1, the vancomycin-resistant *E. faecalis (*strain ATCC 700802 / V583) and the
203 chloramphenicol-resistant *P. aeruginosa* (PAO1). The media applied was LB (Alpha Biosciences,
204 Baltimore MD US).

205 2.2.2. Samples

206 The 188 inlet wastewater samples (40-50 ml) were collected in time-series of 2-4 days during July
207 and August 2017, from 48 Danish wastewater treatment facilities geographically distributed in both
208 rural and urban areas on Zealand, Funen and in Jutland. Upon receipt, the samples were centrifuged
209 (9000 x g, 4 °C, 10 min), the supernatant filtered (0.45 µm) and then stored in aliquots at -20°C. The
210 organic waste samples were collected from four different Danish facilities in February, May and
211 November 2017. The 12 soil samples (~5 g) were collected in Roskilde municipality, Denmark, in
212 March 2019. Prior to screening the soil was suspended in 5 ml LB and slowly and continuously
213 inverted for 1h at room temperature, then the samples were centrifuged (9000 x g, 5 min) and the
214 supernatant filtrated (0.45 µm). Refer to Table S1 for a list of all samples and facilities.

215 2.2.3. DNA extraction, library preparation and sequencing

216 Phage DNA extractions were performed by an initial DNase treatment, 1 U of DNase 1 (New
217 England Biolabs, Ipswich, MA US) per ~100 µl lysate (37°C, 30 min, inactivated by 10 µl 50 mM
218 EDTA), followed by addition of 3 U Proteinase K (A & A Biotechnology, Gdynia, Poland) and 10%
219 (v/v) SDS solution (55°C, 30 min, inactivated by 70°C, 10 min). The extracted DNA was then purified
220 in the well-plate format using the ZR-96 Clean and Concentrator kit (Zymo research, Irvine, CA US),
221 following manufacturer's instructions and eluting in 6 µl of the supplied elution buffer. Sequencing
222 libraries were built according to manufacturer's instructions with minor modifications as described
223 in Kot *et al.*, (2014) [17] using the Nextera® XT DNA kit (Illumina, San Diego, CA USA), the libraries
224 were sequenced as paired-end reads on Illumina NextSeq platform with the Mid Output Kit v2 (300
225 cycles).

226 2.2.4. Assembly, annotation, identification and phylogenomic analysis

227 The obtained reads were trimmed and assembled in CLC Genomics Workbench 10.1.1. (CLC
228 BIO, DK), overlapping reads were merged with the following settings: mismatch cost: 2, minimum
229 score: 15, gap cost: 3 and maximum unaligned end mismatches: 0, and then assembled *de novo*.
230 Additional assemblies were constructed using SPAdes 3.12.0 [19]. Gene prediction and annotation
231 was performed using a customized RASTtk version 2.0 [20] workflow with GeneMark [21], with
232 manual curation and verification using BLASTP [22], HHpred [23] and Pfam version 32.0 [24], or *de*
233 *novo* annotated using VIGA version 0.11.0 [25] based on DIAMOND searches (RefSeq Viral protein
234 database) and HMMer searches (pVOG HMM database). NT similarity was determined as
235 percentage query cover multiplied by percentage NT identity. Novel phages were categorised
236 according to ICTV taxonomy. The criterion of 95% DNA sequence similarity for demarcation of
237 species was applied to identify novel species representatives and to determine species uniqueness
238 within the dataset. All unique phage genomes were deposited in GenBank (Table 1). All genomes
239 were assessed for antibiotic resistance genes (ARGs) and bacterial virulence genes using ResFinder
240 3.1 [26], [27] and VirulenceFinder 2.0 [28]. NT and amino acid (AA) similarities were calculated using
241 tools recommended by the ICTV [29], i.e. BLAST [22] for identification of closest relatives (BLASTn
242 when possible, discontinuous megaBLAST (word size 16) for larger genomes) and Gegenees version
243 2.2.1 [30] for assessing phylogenetic distances of multiple genomes, for both NTs (BLASTn algorithm)
244 and AAs (tBLASTx algorithm) a fragment size of 200 bp and step size 100 bp was applied.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

245 Evolutionary analyses for phylogenetic trees were conducted in MEGA7 version 2.1 (default settings)
246 [31]. These were based on the large terminase subunit *terL*, a gene commonly applied for phylogenetic
247 analysis [32], [33] and on the DNA encapsidation gene *gpA* for the <13kb alleged *Podoviridae*. The NT
248 sequences were aligned by MUSCLE [34] and the evolutionary history inferred by the Maximum
249 Likelihood method based on the Tamura-Nei model [35]. The tree with the highest log likelihood is
250 shown manually curated by adding color-codes and identifiers in Inkscape version 0.92.2. The R
251 package iNEXT [36], [37] in R studio version 1.1.456 [38] was used for rarefaction analyses, species
252 diversity (q = 0, datatype: incidence_raw), extrapolation hereof (estimadeD) and estimation of sample
253 coverage. Additional graphs were prepared in Excel version 16.31.

254 ## 3. Results

255 ### *3.1. Screening efficiency and resolution*

256 Across all five screening between 3% (n = 5 of 188) and 81% (n = 153 of 188) of samples yielded
257 clearing zones plausibly due to lysis by phages, the majority of these also gave rise to the
258 identification of phages (Table 1). However, in some cases the DNA extraction was unsatisfactory or
259 the sequencing failed. Between 61% (*E. coli* screening n = 94 of 153) and 82% (*S. enterica* SV screening
260 n = 42 of 51) of clearing zones were successfully sequenced i.e. yielded reads assembling to phage
261 contigs with an average coverage > x 20 (Table 1). Regardless of host, a single phage was identified
262 from the vast majority of sequenced samples (64-100% per screening), although in some instances
263 two (0-29% per screening), three (0-6% per screening) or four (0-1% per screening) phages were
264 identified from a single sample (Figure 2a). The *Escherichia* phages were the most numerous (136
265 phages from 94 wells), they were more frequently (34 samples) isolated as more than one phage per
266 sample and the only ones to be four phages in a sample [39] (Figure 2a, Table 1). The number of
267 phages per sample did not differ considerably between phages of *S. enterica* (123 phages form 102
268 wells) and *P. aeruginosa* (43 phages from 38 wastewater wells), while only four phages lytic to *E.*
269 *faecalis* were identified in four separate samples (Figure 2a). *Escherichia* phages were identified in
270 samples from 43 different facilities out of the 48, *Enterococcus* phages in samples from 4 facilities and
271 *Salmonella* phages in samples from 22 of the 48 facilities included in these screenings. In the *P.*
272 *aeruginosa* phage screening, phages were identified in wastewater samples from 95% of the 21
273 facilities included, (20 out of 21). Furthermore, *P. aeruginosa* phages were identified in three of the
274 four organic waste samples, but in none of the 8 soil samples (Figure 2c).

275 **Table 1.** Number of samples screened, clearing zones detected and phages identified in the five HiTS screenings,
276 using *S. enterica, E. coli, E. faecalis* or *P. aeruginosa* as host. *Escherichia* phage data from Olsen *et al.*, (2020) [39].

Host	Sample (ml)	Samples (n)	Clearing zones (n)	Sequenced lysates[1] (n)	Phages (n)	Unique species[2] (n)	Novel species[3] (n)
S. enterica	0.5 SV	188	51	42	47	33	28
S. enterica	1.5 LV	188	74	60	76	45 (26[4])	38 (24[4])
E. coli	0.5 SV	188	153	94	136	104	91
E. faecalis	0.5 SV	188	5	4	4	4	3
P. aeruginosa	1.5 LV	94[5]	48	38	46	22	8

277 [1]Includes all lysates for which sequencing yielded reads assembling to contigs with a coverage > x 20. [2]Phages
278 with ≤ 95 % nucleotide similarity to the other phages in this dataset. [3]Phages with ≤ 95 % similarity to the other
279 phages in this dataset and those deposited in the NCBI database. [4]Excluding the phages with >95 % nucleotide
280 similarity to phages in the *S. enterica* SV screening. [5]82 wastewater samples, 8 soil samples and 4 organic waste
281 samples.

282 Of the 136 *Escherichia* phages, the majority (76%) represent unique species [39]. The many
283 *Salmonella* and *P. aeruginosa* phages are more homogeneous. The two *S. enterica* phage screenings
284 identified 123 phages. Out of 47 SV phages 14 were shown to have >95% NT similarity with other SV
285 phages, while 31 of the 76 LV phages were shown to have >95% NT similarity with other LV phages

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and an additional 19 of the LV phages had >95% NT similarity with SV phages. Hence, a total of 59 (48%) *Salmonella* phages of distinct species are identified. Similarly, of the 46 *P. aeruginosa* phages 22 (48%) are unique, while all four *E. faecalis* phages represent distinctive species (Table 1).

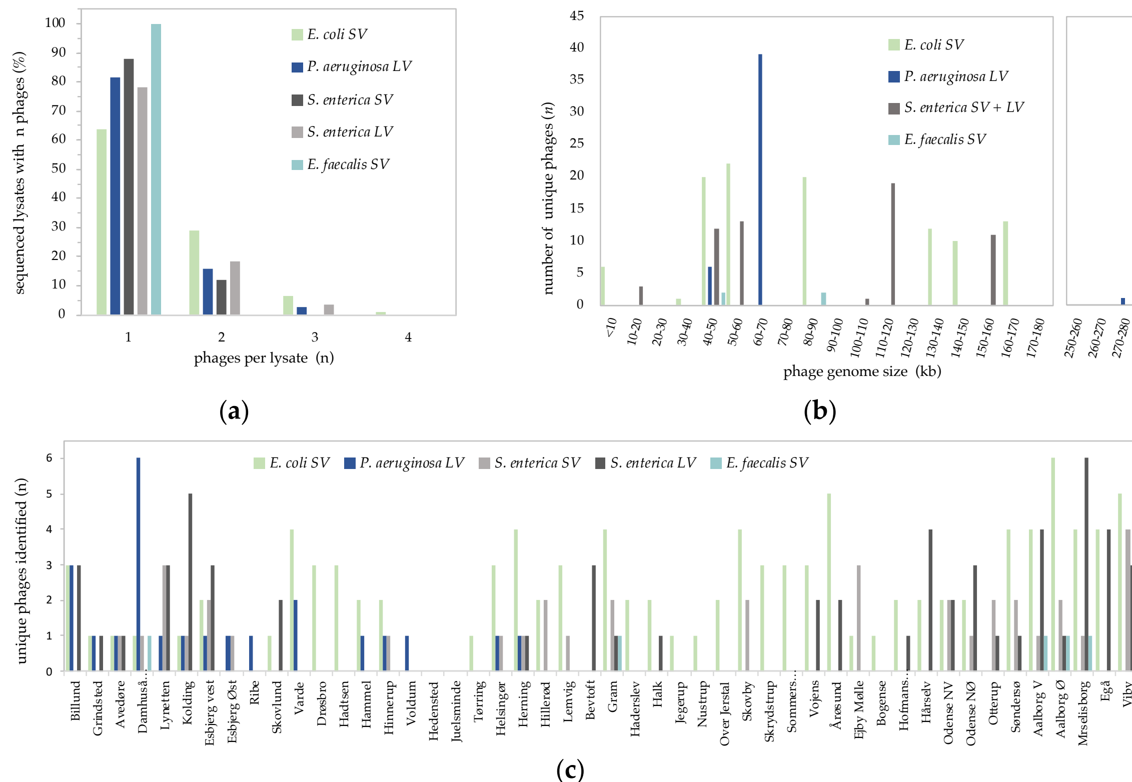


Figure 2 (**a**) Number of phages per lysate (n) (x-axis) as occurring in percentage of all sequenced lysates (y-axis), presented according to individual screenings. (**b**). Average genome size (kb) of Gegenees based clusters, of unique phage species (<95% nucleotide similarity to other phages in the dataset) organized by host, from all screenings. (**c**) Distribution of all 309 phages identified organised per facility, only the first 21 facilities were included in the *Pseudomonas aeruginosa* screening.

3.2. Novelty and diversity of HiTS-phages identifed compared to NCBI

The phages identified cover an impressive wide range of genome sizes (Figure 2b, 4), GC contents and predicted morphologies, representing five different families; the *Ackermannviridae, Myoviridae, Podoviridae* and *Siphoviridae* of the order *Caudovirales* and also the non-tailed *Microviridae* (Table 2) [39]. The *Escherichia* phages are indeed remarkably numerous and diverse and are consequently described separately in Olsen *et. al.* (2020) [39]. In summary, disregarding the jumbo phage Pseudomonas phage fnug (278.9 kb), the *Escherichia* phages cover the largest size range (5.3-170.8 kb) and have an impressive GC content span (35.3-60.0%) [39]. Members of the new family *Ackermannviridae,* were only detected among the *Salmonella* phages, just as members of the *Microviridae,* a family of small single stranded DNA phages, were only observed among the *Escherichia* phages [39]. The *Salmonella* and *P. aeruginosa* phages covered similar GC content spans of 36.9-56.5% and 39.0-59.5%, respectively. The *Salmonella* phage genomes vary in sizes from 11.6 kb (Salmonella phage astrithr) to 159.1 kb (Salmonella phage maane), while the non-jumbo *P. aeruginosa* phages are more uniform having genome sizes of 44.9 kb (Pseudomonas phage clash) to 66.5 kb (Pseudomonas phage shane). The *Enterococcus* phages have genomes of 39.7-85.7 kb, Enterococcus phage heks and nattely, respectively (Figure 2b, Tables S2-S4).

An impressive number (n = 154) of novel phage species have so far been identified with the HiTS method. No less than 67% (n = 91) of the *Escherichia* phages [39], 42% (n = 52) of the *Salmonella* phages, 22% (n = 8) of the *P. aeruginosa* phages and three out of four *Enterococcus* phages represent novel phage species (Table 1, Figure 3). Whereas most of the *Escherichia* (69%) and all of the *P. aeruginosa* phage

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

species representatives have a high NT similarity (>89%) with their closest relatives, a larger proportion of the *Salmonella* phages differ more from their closest relatives as only 54% (n = 42) of the unique *Salmonella* phages species have >89% NT similarity to their closest relative (Figure 3). Two of the *Escherichia* phages and three *Salmonella* phages share <50% NT similarity with published phages. Likewise, two of the *Enterococcus* phages (Figure S3, Table S3) and Salmonella phage Akira (63% NT similarity) are only distantly related to any published phage (62-65% NT similarity) (Figure 3). The 189 unique phages (<95% NT similarity with other phages in the dataset) and their GenBank accession numbers are listed in supplementary materials (Table S2-S5).

Table 2 Predicted morphology of all the phages identified in screenings for *Escherichia, Salmonella, P. aeruginosa* and *Enterococcus* phages, based on taxonomy of closest relative. *Escherichia* phage data from Olsen *et al.*, (2020) [39].

Phage taxonomy	Isolation hosts:	*E. coli*	*S. enterica*	*P. aeruginosa*	*E. faecalis*
Caudovirales; Ackermannviridae		-	21	-	-
Caudovirales; Myoviridae		79	18	39	-
Caudovirales; Siphoviridae		34	76	1	4
Caudovirales; Podoviridae		10	8	6	-
Microviridae; Bullavirinae		8	-	-	-
Unclassified bacterial viruses		5	-	-	-

No virulence factors or ARGs were detected in any of the 189 unique phages. Furthermore, no integrases were identified and though putative recombinases do occur it is uncertain if they are involved in DNA repair or integration.

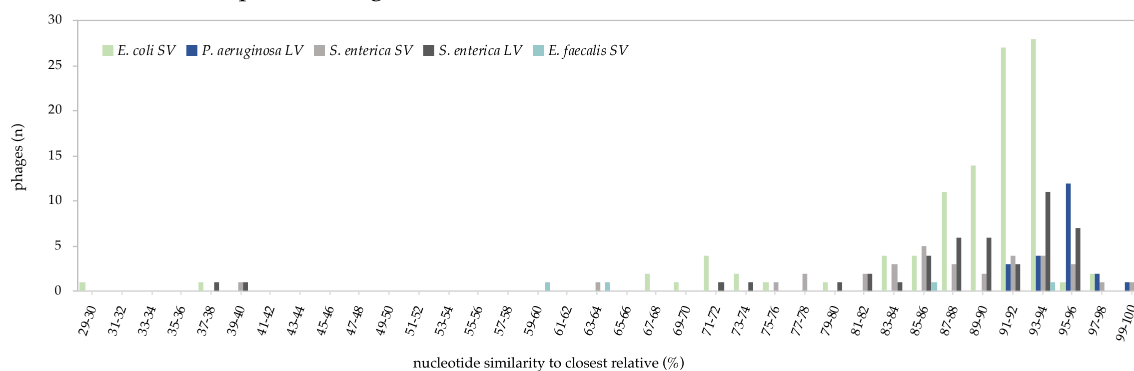


Figure 3 Distribution of nucleotide similarity (%) to closest relative (Blast) of all phages from all five screenings.

3.3 P. aeruginosa and Enterococcus phages from Danish wastewater

The *P. aeruginosa* phages group into two distinct clusters and two singletons (Figure S2). The vast majority (n = 38) of the *P. aeruginosa* phages are closely related (90.9-98.5% NT similarity) to phages of the genus *Pbunavirus* of the family *Myoviridae*, with genome sizes of 60.7-66.8 kb (89-95 CDSs, 54.8-55.7% GC, no tRNAs,). A smaller group of six *P. aeruginosa* phages (44.9-45.3 kb, 68-69 CDSs, 52.1-52.5% GC, 3-4 tRNAs) are closely related (94.1-96.3% NT similarity) to phages of the genus *Bruynoghevirus* of the family *Podoviridae*. The jumbophage fnug is closely related (93% NT similarity) to phages of the genus *Phikvirus*, family *Myoviridae*. While the last Pseudomonas phage Iggy (60.7 kb, 90 CDSs, 56.5% GC, no tRNAs) is closely related (94.6%) to the unclassified *Siphoviridae* Pseudomonas phage PBPA162 (MK816297), none of them share >8% NT similarity with any other published phages.

The *Enterococcus* phages are all predicted to have *Siphoviridae* morphology, but divide into two distinct clusters with NT inter-Gegenees scores of 0 (Figure S3). Phages heks and Nonaheksakonda (39.7-41.9 kb, 64-74 CDSs, 34.6-35.0% GC, no tRNAs) are related to efquatroviruses, but with only

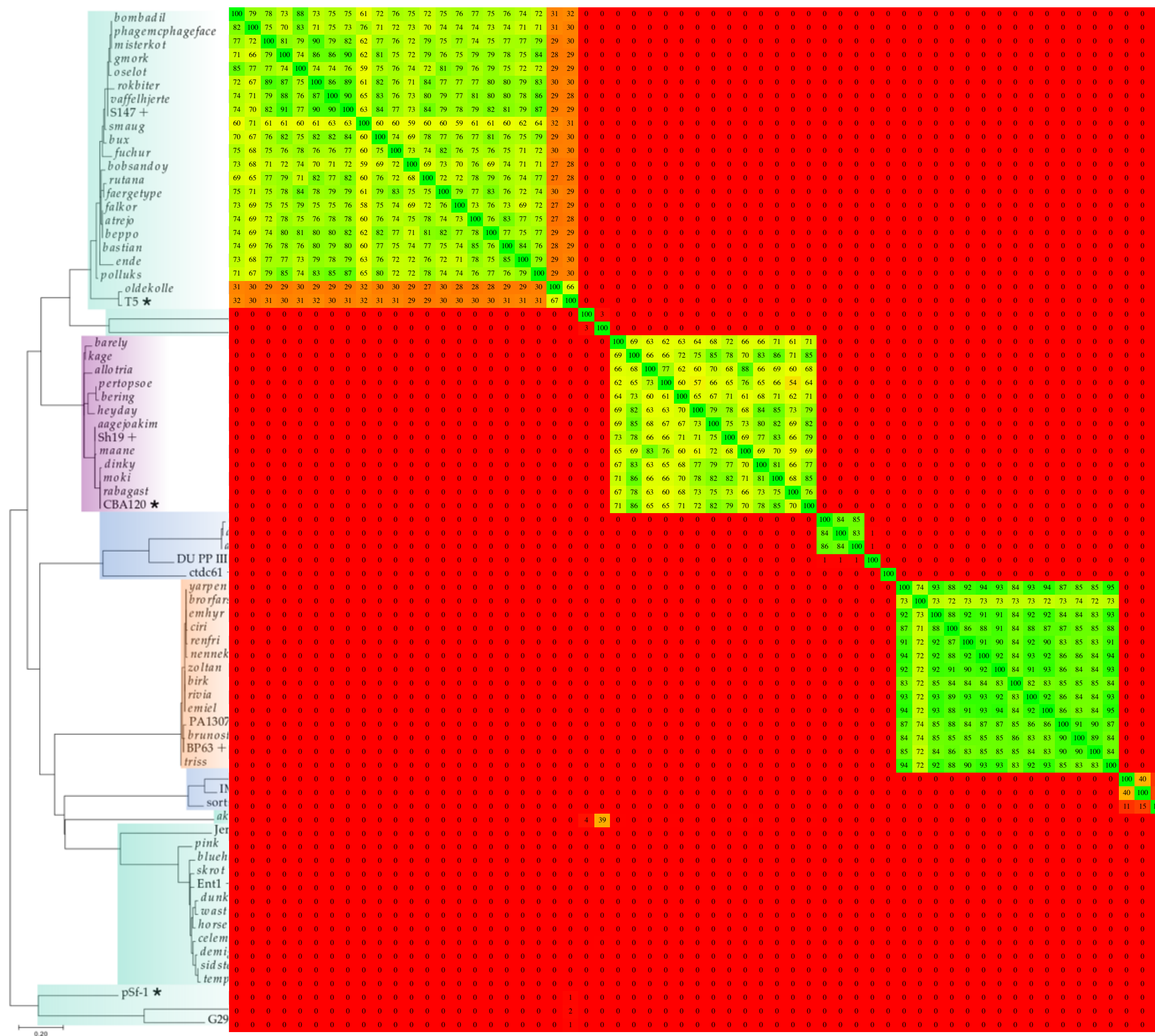
bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

346 59% NT similarity. The other two (85.3-85.7 kb, 131-134 CDSs, 30.2-30.3% GC, 1 tRNA) are more
347 closely related (87-96% NT similarity) to unclassified *Siphoviridae* (Figure 2b, Tables S2-S5).

348 *3.4 Salmonella phages from Danish wastewater*

349 Based on NT similarity with closest relatives, 59 distinctive species of *Salmonella* phages were
350 identified, of which 52 represent novel species. Estimations based on both the SV and LV screenings
351 predicts species richness of easily culturable phages lytic to *S. enterica subsp. enterica* serovar
352 Enteritidis PT1 in Danish wastewater to be nearby 80 (Figure S4), while Shannon diversity estimates
353 68 and 61 and Simpson diversity 51 and 38, for the SV and LV screenings, respectively (Figure S4).
354 The estimates are however expected to be subject to large prediction bias due to the relatively small
355 reference sample size, and a 95% confidence interval suggests a range for all diversity indices of 26-
356 173. Sample completeness is estimated to be achieved at ~1300 samples for a SV screening and at ~800
357 samples for a LV screening (Figure S4). The HiTS *Salmonella* phages belong to at least four different
358 families, *Ackermannviridae*, *Myoviridae*, *Podoviridae* and *Siphoviridae*, covering a wide range of genome
359 sizes and GC contents (Table S5). They group into five clusters and three singletons (inter-Gegenees
360 scores = 0) corresponding to their proposed taxonomy, excluding phage Akira (Figure 4). In spite of
361 a NT similarity of 63% and a Gegenees score of 38-39 with its closest relative the unclassified *roufvirus*
362 Salmonella virus KFS_SE2 (MK112901) Akira does not group with neither KFS_SE2 or the type
363 species of *roufvirus* Aeromonas phage pIS4-A (NC_042037) in the phylogenetic tree, resulting in a
364 peculiar pattern in the Gegenees analysis (Figure 4). Furthermore, Akira shares limited NT similarity
365 (39%) with pIS4-A and a Gegenees score of only 3-4 (Figure 4, Table 2).
366
367 All nine novel phages (143-159 kb, 187-209 CDSs, 44-45% GC, 3-4 tRNAs) of the *Ackermannviridae* are
368 related (78-99% NT similarity) to phages of the genus *Kuttervirus*, subfamily *Cvivirinae*, defined by an
369 average genome size of 158.1 kb, with and average GC content of 44.5% and averagely coding for 201
370 proteins and 4.3 tRNAs. These nine phages have an average NT intra-Gegenees score of just 60, but
371 are more similar to one another and also to the type species of *Kuttervirus,* Escherichia virus CBA120
372 (JN593240) when comparing AAs (Gegenees score = 88-97).
373 Most of the *Salmonella* phages are predicted to belong to *Siphoviridae* (n = 76, 32 species, 29 novel
374 species). Twenty of the unique *Siphoviridae* phages (105-115 kb, 151-171 CDSs, 39.3-40.1 GC, 23-31
375 tRNAs) are related (71-95% NT similarity) to phages of the genus *Tequintavirus*, the T5 like phages.
376 However, whereas phage oldekolle has low NT Gegenees scores with the rest (n = 27-32) and is closely
377 related to T5 (93% NT similarity), the remainder appear more distantly related to T5. Even though,
378 these phages cluster together, they have relatively low NT inter-Gegenees scores (n = 59-90) and they
379 also differ from published phage genomes by 6-29% NT similarity (Figure 4, S5, Table S5). An
380 additional ten unique *Siphoviridae* phages (41-44 kb, 56-70 CDSs, 50% GC, 0-1 tRNA) were found to
381 be closely related to jerseyviruses, subfamily *Guernseyvirinae*. This genus is defined by genome sizes
382 of 40-44 kb, comparable morphology and a shared DNA identity of ≥60% and >68% protein content
383 [40]. The wastewater jerseyviruses-like phages are a heterogeneous group with varying NT intra-
384 Gegenees scores of 38-91, yet the AA intra-Gegenees scores are all >69. However, the Gegenees NT
385 scores between the novel phages and the type species Salmonella phage Jersey (NC_021777) are all
386 <33 though the AA scores are 65-71. The novel jerseyviruses-like phages are relatively abundant in
387 the Danish wastewater samples and most of them were identified in several samples from different
388 treatment facilities. Phages with >95% NT similarity with phage wast were found eight times in
389 samples from five distinct facilities and phages with >95% NT similarity with phage demigod as
390 many as 12 times, in samples from six different facilities (Table S5).
391
392 The last *Siphoviridae* within the *Salmonella* phages is the novel phage slyngel, related (92% NT
393 similarity) to Escherichia phage vB_EcoS_G29-2 (MK373798) an unclassified *Hanrivervirus*, subfamily
394 *Tunavirinae*. Furthermore, slyngel has 88% NT similarity with the type species Shigella phage pSf-1
395 (NC_021331), with which a more distant relationship is also suggested by the phylogenetic analysis
396 (Figure 4, S6).

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



397 **Figure 4** Phylogenetic tree (Maximum log Likelihood: -4176.67, based on large terminase subunit or for the <13 kb
398 *Podoviridae* the DNA encapsidation protein, scalebar: substitutions per site) and phylogenomic nucleotide distances
399 (Gegenees, BLASTn: fragment size: 200, step size: 100, threshold: 0%). Predicted morphology is indicated by colorbars,
400 *Myoviridae*, *Siphoviridae*, *Podoviridae*, *Ackermannviridae*, novel phages from this study are in italics, while close
401 relatives are denoted by + and type species by ★

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

402 Twelve of the *Salmonella* phages are based on NT similarity of the *Myoviridae* and though their
403 closest relatives are all unclassified *Myoviridae*, they constitute the most homogeneous group. They
404 have comparable genomes (52-53 kb, 67-73 CDSs, 45.7-46% GC, no tRNAs) and cluster together in
405 both the Gegenees analyses and in the phylogenetic tree. Though phage brorfarstad has slightly lower
406 NT Gegenees scores (n = 72-74), than the rest, only a minor difference can be observed in the AA
407 Gegenees analysis (Figure 4, S5, Table 3).
408
409 The eight *Podoviridae* divide into a cluster of three phage species representatives (astrid, assan
410 and astrithr) with comparable (83-84 NT Gegenees score) small low GC genomes (11.6-11.7 kb, 15
411 CDSs, 39.7-39.8% GC, no tRNASs) and the singleton lumpael. Phages with >95% NT similarity with
412 phage Lumpael were observed in five samples from five distinct facilitates. Lumpael (41.1 kb, 58
413 CDSs, 59.9% GC, no tRNAs) has the highest GC content observed and shares only 76% NT similarity
414 with its closest relative Enterobacteria phage IME_EC2 (KF591601). Astrid, assan and astrithr all
415 share <40% NT similarity with their closest relative Pectobacterium phage DU_PP_III (MF979562),
416 and though they share NT Gegenees scores of only 0-1 and AA of 45, they do form a monophyletic
417 clade (Figure 4, S5).

418 **4. Discussion**

419 When it comes to phage isolation three aspects are key; titre, sterility and purity. These features are
420 a prerequisite for any phage work, regardless of aim. However, purity and sterility are not the main
421 focus in a screening such as the HiTS method. This method does not intend to provide a final PT or
422 biocontrol product, but instead offers crude isolation of highly diverse phages in sufficient titres. This
423 is to provide a fast and crude acquisition of numerous and diverse phages and thereby a basis for
424 further phage isolation and establishment or expansion of phage libraries. If a phage of interest is in
425 a mixed lysate, sequencing enables primer design for PCR verification when isolating the
426 individually plaquing phages.
427 The method is unique in its ability to identify numerous assorted phages while also facilitating
428 crude isolation in a very short time span. This makes investigations of phage diversity easy and
429 possible. The capacity of the HiTS method to uncover diverse phages is clearly illustrated by the
430 impressive findings when screening for phages of PT relevance in wastewater, especially those
431 infecting *E. coli* and *S. enterica.*, as presented in Olsen *et al.* (2020) [39] and in this study, respectively.
432 Even for less abundant phages diversity and novelty was uncovered, the *P. aeruginosa* phages
433 represent three distinct families, and eight are novel phage species representatives, while all the
434 *Enterococcus* phages are of the *Siphoviridae* two of them have limited NT similarity (<60%) with
435 published phages. The reported five screenings yielded no less than 331 potential hits in the form of
436 clearing zones resulting in the identification of 154 novel phage species. Furthermore, none of these
437 phages code for known virulence factors or ARGs and none appear to be lysogenic, making them all
438 potential candidates for PT and biocontrol applications.
439 Unlike metagenomic sequencing approaches, this method provides actual phages with a direct
440 link to the pathogenic host or any host in interest. The HiTS method does not reveal the diversity of
441 individual samples; hence many phages remain undetected, especially those which are not easily
442 grown under laboratory conditions. The HiTS method is a competition-based method and clearly
443 selects for lytic phages with traits preferable in PT and biocontrol applications i.e. a high burst size
444 and a short latency period. Consequently, the HiTS method enables the capture of the most prevalent
445 phage(s) of the day in any sample. Thus, when screening numerous distinct samples, it provides an
446 estimate of species richness of this type of phages in the given sample matrix. Accordingly, the species
447 richness of easily culturable phages, presumably with high burst-sizes and short latency times, lytic
448 to the specific strains of *Escherichia* and *Salmonella* phages in Danish wastewater was estimated to be
449 at least in the range of 160-420 and 49-173, respectively (Figure S4) [39]. This is likely an
450 underestimation considering the relatively small sample sizes and the inherent bias in the method to
451 only isolate a single or a few phages per sample combined with the many plaque-forming lysates for
452 which DNA-extraction or sequencing was unsuccessful. If the aim of screening is to isolate phages

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

453 with PT or biocontrol potential or phages that are easy to study under laboratory-conditions, the
454 targeting of lytic and highly reproductive phages is indeed an advantage. However, if the aim is to
455 disclose true diversity or detect more difficult to culture specimen, other methods such as plaquing
456 without amplification may be superior. Metagenomic sequencing approaches are constantly refined
457 and now offers high detection levels of phageomes [41], but phages of interest detected may be near
458 impossible to isolate *in vitro*. The key advantage of the HiTS method is indeed that it offers both
459 identification through sequencing and also provides physical isolates of all phages targeting the
460 specific host-species used as bait. Consequently, interesting discoveries such as rare and novel phages
461 or the presence of remarkable genes with unexpected or desired functions can be investigated
462 following a final isolation. It should however be noted that it is not recommended to sequence the
463 lysate giving rise to plaques or clearing zones, while harvesting phages from the plaques, as was done
464 by the authors in the SV screenings. This approach may result in sequencing of phages in lysate not
465 present in the harvested soft-agar.
466 A 96-well setup carries a risk of cross-contamination, however the use of pierceable sealing-tape
467 as recommended in the HiTS method, reduces this risk. The presence of a negative amplification
468 control in each spot-test (agar-plates A and B) provides an indication of potential cross-
469 contamination. No plaquing was observed in negative amplification controls in any of the screenings.
470 Furthermore, the chequered-pattern with empty wells between all lysates used when performing the
471 spot-tests ensures that in the case of improper sterilisation, still no phages will be transferred to other
472 wells in use during spotting with a 96-pin replicator, as opposite patterns are present in well-plate A
473 and B. Finally, if the sterilisation of the 96-pin replicator is insufficient any contaminating phages will
474 plaque in between purposely spotted phages. This was not observed in any of the screenings.
475 A high number of distinct samples, as required by the HiTS method, may be cumbersome to
476 collect and prepare, but once they are collected, they can be aliquoted, stored (-20°C) and used for
477 numerous screenings of different target bacteria, as only very small sample volumes (0.5-1.5 ml) are
478 required. The small sample-volumes also permits for samples to be collected by sending out
479 collection-kits and having them returned by mail or carrier, provided that the applicable law allows
480 it. Any sample matrix with high quantities of the target host is applicable. Furthermore, the suitability
481 of time-series of wastewater-samples eases the sample collection and makes the screening method
482 more feasible as it limits the number of distinct sampling sites. In this study, no *Enterococcus*, only
483 two *P. aeruginosa* (9%), five *Salmonella* (8.5%) and nine *Escherichia* phage species (8.6%) [39] were
484 detected more than once in samples from the same facility in any distinct screening. Wastewater
485 treatment plants receive inlet wastewater in a constant yet changing flow thus the presence of diverse
486 phages can be expected to fluctuate and be interchangeable.
487 Sequencing is continuously getting cheaper [42], and even though this is the major expenditure
488 of the HiTS method, it is economically feasible. Spending weeks and months on thoroughly isolating
489 hundreds of phages is also a costly affair in regards of time and workhours. And still, also by
490 individual isolation resulting phages may end up being similar specimen.
491 With the HiTS method phage libraries can be build and sequenced after just four consecutive
492 days of sample processing. The fast turnaround is of particular importance when screening for
493 phages for PT, but not all phages are suitable for PT. Lysogenic phages should be avoided as they do
494 not necessarily lyse their hosts and may also increase virulence of their hosts by lysogenic conversion
495 [43]. Some phages, also those with a lytic lifestyle, code for genes with unwanted genetic traits such
496 as toxins, superantigens, intracellular survival/host cell attachments proteins or ARGs which can be
497 spread to bacterial communities through transduction [44], [45]. This is especially relevant to consider
498 when isolating phages from wastewater, as treatments plants can be considered hotspots for ARGs
499 [45]. Fortunately, ARGs and other unwanted genetic traits can, for a large part, be deduced by genetic
500 analyses and thus phages coding for them can with decent confidence be excluded. Hence, the HiTS
501 method allows selection of new candidate phages after a few weeks of screening, sequencing and
502 analysing. The ability of the candidate phages to infect the target host is already verified and the
503 absence of undesired genetic traits confirmed, consequently the phages are now ready for
504 experimental validation and final isolation, if required.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

505 In conclusion, the HiTS method presented here has the potential to efficiently detect the diversity
506 of and crudely isolate phages relevant for PT and biocontrol which are abundant in the sample matrix
507 explored. The HiTS method is simple, fast and cost-efficient. It can prove to be a valuable, scalable
508 method in the case of urgent needs for PT suitable phages targeting specific clinical infections. With
509 the HiTS method establishment of sizeable discovery phage banks becomes fast and efficient. Such
510 phage discovery banks could be lifesavers eliminating the need to spend time on isolating new PT
511 phages and would also facilitate important phage taxonomy and ecology studies and can be explored
512 for industrially-relevant biotechnological applications.

513 **Supplementary Materials:** The following are available online at www.mdpi.com/xxx/s1, Figure S1: Illustration
514 of Checker plating, Table S1: List of samples, Table S2: List of unique *P. aeruginosa* phages, Table S3: List of
515 unique *Enterococcus* phages, Table S4: List of unique *Salmonella* phages, Figure
516 S2: Phylogenomic nucleotide distances of the 39 unique *P. aeruginosa* phages, Figure S3:
517 Phylogenomic nucleotide distances of the four unique *Enterococcus* phages, Figure S4: Rarefaction curves and
518 diversity indices, Figure S6: Phylogenomic amino acid distances of the 59 unique *Salmonella* phages.

519 **Author Contributions:** Conceptualization, N. S. O., W. K.: and L. H. H.; methodology, N. S. O., W. K. and L. H.
520 H.; validation, N. S. O., W. K., N. B. H. and L. H. H.; formal analysis, N. S. O.; investigation, N. S. O.; resources,
521 N. S. O. and L. H. H.; data curation, N. S. O.; writing—original draft preparation, N. S. O.; writing—review and
522 editing, N. S. O., W. K., N. B. H. and L. H. H.; visualization, N. S. O.; supervision, N. S. O., W. K. and L. H. H.;
523 project administration, N. S. O., W. K. and L. H. H.; funding acquisition, W. K. and L. H.H.

524 **Funding:** This research was funded by Villum Experiment Grant 17595, Aarhus University Research Foundation
525 AUFF Grant E-2015-FLS-7-28 to Witold Kot and Lars Hestbjerg Hansen's Human Frontier Science Program
526 Grant: RGP0024/2018.

527 **Acknowledgments:** This study had not been possible without the much-appreciated contribution by the many
528 collaborating members of the Danish Water and Wastewater Association (DANVA), who kindly supplied us
529 with time-series of wastewater samples from their treatment facilities. A special thanks to the Billund and
530 Grindsted treatment facilities of Billund Vand & Energi, Lynetten, Avedøre and Damhusåen treatment facilities
531 of BIOFOS, Kolding treatment facility of BlueKolding, Esbjerg Øst, Esbjerg Vest, Ribe, Varde og Skovlund
532 treatment facilities of DIN Forsyning, Drøsbro, Hadsten, Hammel, Hinnerup and Voldum treatment facilities of
533 Favrskov Forsyning, Haerning treatment facility of Herning Vand, Hillerød treatment facility of Hillerød
534 Forsyning,Lemvig treatment facility of Lemvig Vand og Spildevand, Bevtoft, Gram, Haderslev, Halk, Jegerup,
535 Nustrup, Over Jerstal, Skovby, Skrydstrup, Sommersted, Vojens and Årøsund treatment facilities of Provas,
536 Marselisborg, Egå and Viby treatment facilities of Aarhus vand, Hedensted, Juelsminde and Tørring treatment
537 facilities of Hedensted Spildevand, Ejby Mølle, Bogense, Hofmansgave, Hårslev, Nordvest, Nordøst, Otterup
538 and Søndersø treatment facilities of VandCenterSyd, Øst and Vest treatment facilities of Aalborg Forsyning and
539 finally Helsingør treatment facility of Forsyning Helsingør.

540 **Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the
541 study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to
542 publish the results.

543 **References**

544 [1] F. L. Gordillo Altamirano and J. J. Barr, "Phage Therapy in the Postantibiotic Era.," *Clin. Microbiol. Rev.*,
545 vol. 32, no. 2, pp. e00066-18, Apr. 2019.

546 [2] J. Y. H. Lee *et al.*, "Global spread of three multidrug-resistant lineages of Staphylococcus epidermidis,"
547 *Nat. Microbiol.*, vol. 3, no. 10, pp. 1175–1185, Oct. 2018.

548 [3] A.-P. Magiorakos *et al.*, "Multidrug-resistant, extensively drug-resistant and pandrug-resistant
549 bacteria: an international expert proposal for interim standard definitions for acquired resistance,"
550 *Clin. Microbiol. Infect.*, vol. 18, no. 3, pp. 268–281, Mar. 2012.

551 [4] D. M. Lin, B. Koskella, and H. C. Lin, "Phage therapy: An alternative to antibiotics in the age of multi-

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

552 drug resistance," *World J. Gastrointest. Pharmacol. Ther.*, vol. 8, no. 3, p. 162, 2017.

553 [5] K. M. Caflisch, G. A. Suh, and R. Patel, "Biological challenges of phage therapy and proposed
554 solutions: a literature review," *Expert Rev. Anti. Infect. Ther.*, vol. 17, no. 12, pp. 1011–1041, Dec. 2019.

555 [6] R. M. Dedrick *et al.*, "Engineered bacteriophages for treatment of a patient with a disseminated drug
556 resistant Mycobacterium abscessus," *Nat. Med.*, vol. 25, no. 5, p. 730, 2019.

557 [7] R. T. Schooley *et al.*, "Development and Use of Personalized Bacteriophage-Based Therapeutic
558 Cocktails To Treat a Patient with a Disseminated Resistant Acinetobacter baumannii Infection.,"
559 *Antimicrob. Agents Chemother.*, vol. 61, no. 10, Oct. 2017.

560 [8] S. Aslam *et al.*, "Early clinical experience of bacteriophage therapy in 3 lung transplant recipients," *Am.*
561 *J. Transplant.*, vol. 19, no. 9, pp. 2631–2639, Sep. 2019.

562 [9] A. B. Monk, C. D. Rees, P. Barrow, S. Hagens, and D. R. Harper, "Bacteriophage applications: Where
563 are we now?," *Letters in Applied Microbiology*, vol. 51, no. 4. pp. 363–369, 2010.

564 [10] A. B. Carstens, A. M. Djurhuus, W. Kot, and L. H. Hansen, "A novel six-phage cocktail reduces
565 Pectobacterium atrosepticum soft rot infection in potato tubers under simulated storage conditions,"
566 *FEMS Microbiol. Lett.*, vol. 366, no. 9, May 2019.

567 [11] A. Svircev, D. Roach, and A. Castle, "Framing the future with bacteriophages in agriculture," *Viruses*,
568 vol. 10, no. 5. MDPI AG, p. 218, 01-May-2018.

569 [12] H. Brüssow, "Hurdles for Phage Therapy to Become a Reality—An Editorial Comment," *Viruses*, vol.
570 11, no. 6, p. 557, Jun. 2019.

571 [13] A. Gratia, "Des relations numeriques entre bacteries lysogenes et particules de bacteriophage," *Ann.*
572 *Inst. Pasteur*, vol. 57, no. 1936, pp. 652–676, 1936.

573 [14] K. M. Kauffman and M. F. Polz, "Streamlining standard bacteriophage methods for higher
574 throughput," *MethodsX*, vol. 5, pp. 159–172, Jan. 2018.

575 [15] M. Henry *et al.*, " Development of a high throughput assay for indirectly measuring phage growth
576 using the OmniLog TM system ," *Bacteriophage*, vol. 2, no. 3, pp. 159–167, Jul. 2012.

577 [16] G. F. Hatfull, "Mycobacteriophages: Windows into Tuberculosis," *PLoS Pathog.*, vol. 10, no. 3, p.
578 e1003953, Mar. 2014.

579 [17] W. Kot, F. K. Vogensen, S. J. Sørensen, and L. H. Hansen, "DPS – A rapid method for genome
580 sequencing of DNA-containing bacteriophages directly from a single plaque," *J. Virol. Methods*, vol.
581 196, pp. 152–156, Feb. 2014.

582 [18] E. Wood, "Molecular Cloning. A Laboratory Manual," *Biochem. Educ.*, vol. 11, no. 2, p. 82, 2003.

583 [19] A. Bankevich *et al.*, "SPAdes: a new genome assembly algorithm and its applications to single-cell

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

584 sequencing.," *J. Comput. Biol.*, vol. 19, no. 5, pp. 455–77, 2012.

585 [20] T. Brettin *et al.*, "RASTtk: A modular and extensible implementation of the RAST algorithm for
586 building custom annotation pipelines and annotating batches of genomes," *Sci. Rep.*, vol. 5, no. 1, p.
587 8365, Jul. 2015.

588 [21] J. Besemer and M. Borodovsky, "GeneMark: web software for gene finding in prokaryotes, eukaryotes
589 and viruses," *Nucleic Acids Res.*, vol. 33, no. Web Server, pp. W451–W454, Jul. 2005.

590 [22] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *J.*
591 *Mol. Biol.*, vol. 215, no. 3, pp. 403–410, Oct. 1990.

592 [23] J. Soding, A. Biegert, and A. N. Lupas, "The HHpred interactive server for protein homology detection
593 and structure prediction," *Nucleic Acids Res.*, vol. 33, no. Web Server, pp. W244–W248, Jul. 2005.

594 [24] R. D. Finn *et al.*, "The Pfam protein families database: towards a more sustainable future," *Nucleic Acids*
595 *Res.*, vol. 44, no. D1, pp. D279–D285, Jan. 2016.

596 [25] 6 Enrique González-Tortuero1, Thomas David Sean Sutton1, Vimalkumar Velayudhan1, 4 Andrey
597 Nikolaevich Shkoporov1, Lorraine Anne Draper1, Stephen Robert Stockdale1, 2, Reynolds 5 Paul
598 Ross1, 3, Colin Hill1, 3, "VIGA: a sensitive, precise and automatic de novo VIral Genome Annotator,"
599 *bioRxiv*, p. 277509, Mar. 2018.

600 [26] E. Zankari *et al.*, "Identification of acquired antimicrobial resistance genes.," *J. Antimicrob. Chemother.*,
601 vol. 67, no. 11, pp. 2640–4, Nov. 2012.

602 [27] K. A. Kleinheinz, K. G. Joensen, and M. V. Larsen, "Applying the ResFinder and VirulenceFinder web-
603 services for easy identification of acquired antibiotic resistance and E. coli virulence genes in
604 bacteriophage and prophage nucleotide sequences.," *Bacteriophage*, vol. 4, no. 1, p. e27943, Jan. 2014.

605 [28] K. G. Joensen *et al.*, "Real-Time Whole-Genome Sequencing for Routine Typing, Surveillance, and
606 Outbreak Detection of Verotoxigenic Escherichia coli," *J. Clin. Microbiol.*, vol. 52, no. 5, pp. 1501–1510,
607 May 2014.

608 [29] E. Adriaenssens and J. R. Brister, "How to Name and Classify Your Phage: An Informal Guide.,"
609 *Viruses*, vol. 9, no. 4, 2017.

610 [30] J. Ågren, A. Sundström, T. Håfström, and B. Segerman, "Gegenees: Fragmented Alignment of Multiple
611 Genomes for Determining Phylogenomic Distances and Genetic Signatures Unique for Specified
612 Target Groups," *PLoS One*, vol. 7, no. 6, p. e39107, Jun. 2012.

613 [31] S. Kumar, G. Stecher, and K. Tamura, "MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0
614 for Bigger Datasets," *Mol. Biol. Evol.*, vol. 33, no. 7, pp. 1870–1874, Jul. 2016.

615 [32] A. Lopes, P. Tavares, M.-A. Petit, R. Guérois, and S. Zinn-Justin, "Automated classification of tailed
616 bacteriophages according to their neck organization," *BMC Genomics*, vol. 15, no. 1, p. 1027, Nov. 2014.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

617 [33] D. J. Mercanti *et al.*, "Genomic Diversity of Phages Infecting Probiotic Strains of Lactobacillus
618 paracasei," *Appl. Environ. Microbiol.*, vol. 82, no. 1, pp. 95–105, Jan. 2016.

619 [34] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput.,"
620 *Nucleic Acids Res.*, vol. 32, no. 5, pp. 1792–7, 2004.

621 [35] K. Tamura and M. Nei, "Estimation of the number of nucleotide substitutions in the control region of
622 mitochondrial DNA in humans and chimpanzees.," *Mol. Biol. Evol.*, vol. 10, no. 3, pp. 512–26, May
623 1993.

624 [36] T. C. Hsieh, K. H. Ma, and A. Chao, "iNEXT: an R package for rarefaction and extrapolation of species
625 diversity (Hill numbers)," *Methods Ecol. Evol.*, vol. 7, no. 12, pp. 1451–1456, Dec. 2016.

626 [37] A. Chao *et al.*, "Rarefaction and extrapolation with Hill numbers: A framework for sampling and
627 estimation in species diversity studies," *Ecol. Monogr.*, vol. 84, no. 1, pp. 45–67, 2014.

628 [38] RStudio Team, "RStudio: Integrated Development for R." Boston, MA, 2016.

629 [39] N. S. Olsen, W. Kot, L. M. F. Junco, and L. H. Hansen, "Exploring the remarkable diversity of
630 Escherichia coli Phages in the Danish Wastewater Environment, Including 91 Novel Phage Species,"
631 *bioRxiv*, p. 2020.01.19.911818, Jan. 2020.

632 [40] E. M. Adriaenssens *et al.*, "Integration of genomic and proteomic analyses in the classification of the
633 Siphoviridae family," *Virology*, vol. 477, pp. 144–154, Mar. 2015.

634 [41] C. d'Humières *et al.*, "A simple, reproducible and cost-effective procedure to analyse gut phageome:
635 from phage isolation to bioinformatic approach," *Sci. Rep.*, vol. 9, no. 1, pp. 1–13, Dec. 2019.

636 [42] "DNA Sequencing Costs: Data | NHGRI." [Online]. Available: https://www.genome.gov/about-
637 genomics/fact-sheets/DNA-Sequencing-Costs-Data. [Accessed: 23-Feb-2020].

638 [43] H. Brussow, C. Canchaya, and W.-D. Hardt, "Phages and the Evolution of Bacterial Pathogens: from
639 Genomic Rearrangements to Lysogenic Conversion," *Microbiol. Mol. Biol. Rev.*, vol. 68, no. 3, pp. 560–
640 602, Sep. 2004.

641 [44] R. Reindel and C. R. Fiore, "Phage therapy: Considerations and challenges for development," *Clin.*
642 *Infect. Dis.*, vol. 64, no. 11, pp. 1589–1590, 2017.

643 [45] W. Calero-Cáceres, M. Ye, and J. L. Balcázar, "Bacteriophages as Environmental Reservoirs of
644 Antibiotic Resistance," *Trends in Microbiology*. 2019.

645

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figures and Tables for:

A New High-throughput Screening (HiTS) Method for Phages – Enabling Crude Isolation and Fast Identification of Diverse Phages with Therapeutic Potential

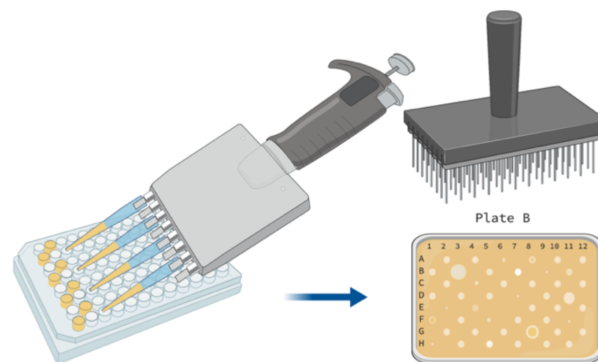


Figure S1 Illustration of chequer plating, removing every second row of pipette-tips enables the easy transfer of lysate to every second well of a well-plate in a chequer patter ensuring a safe distance between clearing zones when plating. Illustration created in BioRender.

Table S1 List of samples used for phage screenings.

Sample type	Company/source	Facility/city	Samples
Wastewater	Billund	Billund	1A, 1B
Wastewater	Billund	Grindsted	2A, 2B
Wastewater	BIOFOS	Avedøre	3A, 3B, 3C, 3D
Wastewater	BIOFOS	Damhusåen	4A, 4B, 4C, 4D
Wastewater	BIOFOS	Lynetten	5A, 5B, 5C, 5D
Wastewater	BlueKolding	Kolding	6A, 6B, 6C, 6D
Wastewater	Din Forsyning	Esbjerg vest	7A, 7B, 7C, 7D
Wastewater	Din Forsyning	Esbjerg Øst	8A, 8B, 8C, 8D
Wastewater	Din Forsyning	Ribe	9A, 9B, 9C, 9D
Wastewater	Din Forsyning	Skovlund	10A, 10B, 10C, 10D
Wastewater	Din Forsyning	Varde	11A, 11B, 11C, 11D
Wastewater	Favrskov Spildevand	Drøsbro	12A, 12B, 12C, 12D
Wastewater	Favrskov Spildevand	Hadsten	13A, 13B, 13C, 13D
Wastewater	Favrskov Spildevand	Hammel	14A, 14B, 14C, 14D
Wastewater	Favrskov Spildevand	Hinnerup	15A, 15B, 15C, 15D
Wastewater	Favrskov Spildevand	Voldum	16A, 16B, 16C, 16D
Wastewater	Hedensted	Hedensted	17A, 17B, 17C, 17D
Wastewater	Hedensted	Juelsminde	18A, 18B, 18C, 18D
Wastewater	Hedensted	Tørring	19A, 19B, 19C, 19D

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Wastewater	Helsingør Fors.	Helsingør	20A, 20B, 20C, 20D
Wastewater	Herning Vand	Herning	21A, 21B, 21C, 21D
Wastewater	Hillerød service	Hillerød	22A, 22B, 22C, 22D
Wastewater	Lemvig Fors.	Lemvig	23A, 23B, 23C, 23D
Wastewater	Provas Fors.	Bevtoft	24A, 24B, 24C, 24D
Wastewater	Provas Fors.	Gram	25A, 25B, 25C, 25D
Wastewater	Provas Fors.	Haderslev	26A, 26B, 26C, 26D
Wastewater	Provas Fors.	Halk	27A, 27B, 27C, 27D
Wastewater	Provas Fors.	Jegerup	28A, 28B, 28C, 28D
Wastewater	Provas Fors.	Nustrup	29A, 29B, 29C, 29D
Wastewater	Provas Fors.	Over Jerstal	30A, 30B, 30C, 30D
Wastewater	Provas Fors.	Skovby	31A, 31B, 31C, 31D
Wastewater	Provas Fors.	Skrydstrup	32A, 32B, 32C, 32D
Wastewater	Provas Fors.	Sommersted	33A, 33B, 33C, 33D
Wastewater	Provas Fors.	Vojens	34A, 34B, 34C, 34D
Wastewater	Provas Fors.	Årøsund	35A, 35B, 35C, 35D
Wastewater	VandCenter Syd	Ejby Mølle	36A, 36B, 36C, 36D
Wastewater	VandCenter Syd	Bogense	37A, 37B, 37C, 37D
Wastewater	VandCenter Syd	Hofmansgave	38A, 38B, 38C, 38D
Wastewater	VandCenter Syd	Hårslev	39A, 39B, 39C, 39D
Wastewater	VandCenter Syd	Nordvest	40A, 40B, 40C, 40D
Wastewater	VandCenter Syd	Nordøst	41A, 41B, 41C, 41D
Wastewater	VandCenter Syd	Otterup	42A, 42B, 42C, 42D
Wastewater	VandCenter Syd	Søndersø	43A, 43B, 43C, 43D
Wastewater	Aalborg Kloak	Vest	44A, 44B, 44C, 44D
Wastewater	Aalborg Kloak	Øst	45A, 45B, 45C, 45D
Wastewater	Aarhus Vand	Marselisborg	46A, 46B, 46C, 46D
Wastewater	Aarhus Vand	Egå	47A, 47B, 47C, 47D
Wastewater	Aarhus Vand	Viby	48A, 48B, 48C, 48D
Soil	Pondside	Roskilde	S1
Soil	Next to rotten tree trunk	Roskilde	S2
Soil	Parking lot	Roskilde	S3
Soil	Potted chives	Roskilde	S4
Soil	Lawn	Roskilde	S5
Soil	molehill	Roskilde	S6
Soil	Potted plant (*Pelargonium graveolens*)	Roskilde	S7
Soil	Flowerbed	Roskilde	S8
Organic waste		Svinninge	O1
Organic waste		Glostrup	O2
Organic waste		Holbæk	O3
Organic waste		Holsted	O4

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S2 List of unique *P. aeruginosa* phages i.e. those which differ by >5% from other phages in the dataset, identified in the LV screening of Danish wastewater, soil and organic waste.

phage	genome size (bp)	GC (%)	CDS (n)	tRNAs (n)	Novel[1]	taxonomy	Sample	Accession
clash	44912	52.1	68	3	*	*Podoviridae; Bruynoghevirus*	1A, 2B, 9A	MT119362
otherone	44930	52.1	68	3		*Podoviridae; Bruynoghevirus*	1B	MT119373
oldone	45313	52.5	69	4	*	*Podoviridae; Bruynoghevirus*	4C	MT119371
Iggy	60769	56.5	90	-		*Siphoviridae.*	4D	MN029011
datas	60746	54.8	89	-		*Myoviridae; Pbunavirus*	21B	MT119378
goonie	64599	55.6	94	-		*Myoviridae; Pbunavirus*	8C	MT133561
elmo	65276	55.0	91	-		*Myoviridae; Pbunavirus*	4B	MT119364
willy	65355	55.6	92	-		*Myoviridae; Pbunavirus*	20C	MT133562
billy	65580	54.9	91	-		*Myoviridae; Pbunavirus*	6D	MT133563
steven	65632	54.9	92	-		*Myoviridae; Pbunavirus*	11D	MT119370
antinowhere	65852	55.2	91	-	*	*Myoviridae; Pbunavirus*	2B, 5D	MT119374
misfit	65887	54.9	93	-		*Myoviridae; Pbunavirus*	3B	MT119367
chumba	65852	55.0	92	-		*Myoviridae; Pbunavirus*	3D, 7D, 8D	MT119375
goodold	66226	55.6	95	-		*Myoviridae; Pbunavirus*	6A, 6B, 9A	MT119365
chunk	66231	55.6	93	-		*Myoviridae; Pbunavirus*	1B	MT119376
debbie	66259	55.7	93	-		*Myoviridae; Pbunavirus*	7A, 8B, 10A, 12D, 15C, 18C, 18D, 21C, O1, O2, O3	MT119363
jett	66262	55.6	93	-	*	*Myoviridae; Pbunavirus*	8A, 11A, 13D	MT119366
crassa	66295	55.2	91	-	*	*Myoviridae; Pbunavirus*	4A	MT119377
cory	66395	55.2	91	-	*	*Myoviridae; Pbunavirus*	4C	MT119372
sortsol	66445	55.7	91	-	*	*Myoviridae; Pbunavirus*	4D, 13B, 19A	MT119369
shane	66467	55.6	95	-		*Myoviridae; Pbunavirus*	14D, 16B	MT119368
fnug	278899	36.9	360	6	*	*Myoviridae; Phikzvirus*	1B	MT133560

[1]Novel phage species classified by a 95% nucleotide similarity demarcation are indicated by an asterisk.

Table S3 List of unique *Enterococcus* phages i.e. those which differ by >5% from other phages in the dataset, identified in the SV screening of Danish wastewater

Phage	genome size (bp)	GC (%)	CDS (n)	tRNAs (n)	Novel[1]	Taxonomy	Sample	Accession
heks	39708	35,0	64	-	*	*Siphoviridae; Efquatrovirus*	4A	MT119359
Nonaheksakonda	41994	34.6	74	-	*	*Siphoviridae; Efquatrovirus*	45D	MK125140
vipetofem	85371	30.3	134	1		*Siphoviridae*	44D	MT119361
nattely	85669	30.2	131	1	*	*Siphoviridae*	46D	MT119360

[1]Novel phage species classified by a 95% nucleotide similarity demarcation are indicated by an asterisk.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S4 List of unique *Salmonella* phages i.e. those which differ by >5% from other phages in the dataset, identified in the SV and LV screenings of Danish wastewater.

phage	genome size (bp)	GC (%)	CDS (n)	tRNAs (n)	Novel[1]	taxonomy	Sample	Accession
astrid	11713	39.8	15	-	*	*Podoviridae; Picovirinae*	7D	MK080312
astrithr	11569	39.7	15	-	*	*Podoviridae; Picovirinae*	7C	MT074429
assan	11578	39.8	15	-	*	*Podoviridae; Picovirinae*	10D	MT074440
lumpael	41126	59.5	58	-	*	*Podoviridae*	3A, 15D, 44A, 46B, 48A	MK125141
wast	40748	49.8	59	1	*	*Siphoviridae; Guernseyvirinae; Jerseyvirus*	28C, 31B, 31C, 31D, 32C, 34B, 34D, 36A	MT074451
sidste	42124	50.0	56	-	*	*Siphoviridae;*	35B	MT074481
templet	41597	49.9	58	1	*	*Siphoviridae; Guernseyvirinae; Jerseyvirus*	6B, 7A, 10B, 10D, 36B	MT074462
blauehaus	41993	49.7	61	-	*	*Siphoviridae; Guernseyvirinae; Jerseyvirus*	47B, 48B, 48C, 48D	MT074473
Horsemountain	42375	49.8	58	1	*	*Siphoviridae; Guernseyvirinae; Jerseyvirus*	21A, 21D, 22A, 22C, 45D	MT074482
skrot	42879	49.7	65	-	*	*Siphoviridae; Guernseyvirinae; Jerseyvirus*	34C	MT074483
dunkel	43985	49.7	64	-		*Siphoviridae; Guernseyvirinae; Jerseyvirus*	8D	MT074432
demigod	43605	50.1	56	1	*	*Siphoviridae; Guernseyvirinae; Jerseyvirus*	2A, 3A, 3C, 3D, 4A, 4C, 2 x 4D, 25D, 26B, 40A, 40D, 41A, 41D, 45C	MT074431
celemicas	43193	49.8	66	1	*	*Siphoviridae; Guernseyvirinae; Jerseyvirus*	2B	MT074484
pink	43203	49.7	70	-	*	*Siphoviridae; Guernseyvirinae; Jerseyvirus*	6D, 7C, 7D, 19D, 20C, 21B	MT074430
akira	45367	46.0	85	1	*	*Siphoviridae; Roufvirus*	21B	MK599416
slyngel	51048	44.0	80	-	*	*Siphoviridae; Tunavirinae; Hanrivervirus*	36A	MT074433
emiel	52333	45.9	67	-	*	*Myoviridae*	6A, 24A	MT074434
brorfarstad	52392	45.7	72	-	*	*Myoviridae*	44D	MT074435
yarpen	52464	46.0	69	-	*	*Myoviridae*	1A	MT074436
rivia	52587	45.9	70	-	*	*Myoviridae*	6B, 47B,	MT074438
nenneke	52576	45.9	71	-	*	*Myoviridae*	44A, 46B	MT074437
birk	52767	45.8	73	-	*	*Myoviridae*	5C	MT074439
brunost	52821	46.0	72	-		*Myoviridae*	5A, 10C,	MT074441
ciri	52916	45.9	71	-	*	*Myoviridae*	27A	MT074442
zoltan	52933	45.9	71	-	*	*Myoviridae*	41D	MT074443

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

emhyr	52948	45.9	71	-	*	*Myoviridae*	35C	MT074444
renfri	52952	45.9	71	-	*	*Myoviridae*	40B	MT074445
triss	53333	45.9	72	-	*	*Myoviridae*	39D	MT074446
bombadil	109539	40.0	163	28	*	*Siphoviridae; Tequintavirus*	38D	MT074449
oldekolle	109372	39.3	157	23	*	*Siphoviridae; Tequintavirus*	20A	MT074448
phagemcphageface	105405	39.9	151	25	*	*Siphoviridae; Tequintavirus*	42D	MT074447
ende	111297	40.0	164	26	*	*Siphoviridae; Tequintavirus*	1B	MT074454
beppo	111322	40.1	167	27	*	*Siphoviridae; Tequintavirus*	46D, 48A	MT074455
rutana	114544	39.9	168	29	*	*Siphoviridae; Tequintavirus*	46B, 48A	MT074468
gmork	113259	39.9	166	29	*	*Siphoviridae; Tequintavirus*	43B, 47A	MT074463
vaffelhjerte	110831	40.0	160	28		*Siphoviridae; Tequintavirus*	6D, 24A, 47C	MT074452
polluks	111599	40.0	162	27	*	*Siphoviridae; Tequintavirus*	40D	MT074456
oselot	113503	39.9	171	28	*	*Siphoviridae; Tequintavirus*	46B	MT074465
faergetype	110183	40.0	161	25		*Siphoviridae; Tequintavirus*	42B	MT074450
bastian	112477	40.0	168	29	*	*Siphoviridae; Tequintavirus*	45A, 46A	MT074459
rokbiter	112006	39.9	163	29	*	*Siphoviridae; Tequintavirus*	39B	MT074457
fuchur	112117	40.0	166	31	*	*Siphoviridae; Tequintavirus*	39C, 39D	MT074458
bux	112486	40.0	160	29	*	*Siphoviridae; Tequintavirus*	46C	MT074460
falkor	114211	39.8	165	27	*	*Siphoviridae; Tequintavirus*	1A, 3A	MT074467
smaug	112590	39.8	161	24	*	*Siphoviridae; Tequintavirus*	42A	MT074461
atrejo	114059	40.0	167	29	*	*Siphoviridae; Tequintavirus*	44A	MT074466
misterkot	110937	39.8	159	28	*	*Siphoviridae; Tequintavirus*	40C	MT074453
bobsandoy	113461	39.9	165	29	*	*Siphoviridae; Tequintavirus*	4B	MT074464
pertopsoe	158905	45.3	202	3	*	*Ackermannviridae; Cvivirinae; Kuttervirus*	8C, 25A, 25C	MT074479
dinky	157254	44.8	204	3	*	*Ackermannviridae; Cvivirinae; Kuttervirus*	41A, 2 x 41D	MT074475
heyday	144232	44.5	187	3	*	*Ackermannviridae; Cvivirinae; Kuttervirus*	7D, 8B, 25B	MT074470
bering	156130	44.7	203	4	*	*Ackermannviridae; Cvivirinae; Kuttervirus*	23B	MT074472
barely	157920	44.6	197	3	*	*Ackermannviridae; Cvivirinae; Kuttervirus*	7A	MT074477
rabagast	143249	44.5	188	3		*Ackermannviridae; Cvivirinae; Kuttervirus*	5A, 5B	MT074469
kage	157658	44.5	205	3		*Ackermannviridae; Cvivirinae; Kuttervirus*	36C, 36D	MT074476
allotria	151015	45.1	202	4	*	*Ackermannviridae; Cvivirinae; Kuttervirus*	5C	MT074471
moki	158230	44.4	209	3	*	*Ackermannviridae; Cvivirinae; Kuttervirus*	44B	MT074478
aagejoakim	156801	44.5	205	-	*	*Ackermannviridae; Cvivirinae; Kuttervirus*	43A, 43D, 40 B	MT074474
maane	159128	45.2	204	4	*	*Ackermannviridae; Cvivirinae; Kuttervirus*	6A	MT074480

[1]Novel phage species classified by a 95% nucleotide similarity demarcation are indicated by an asterisk.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

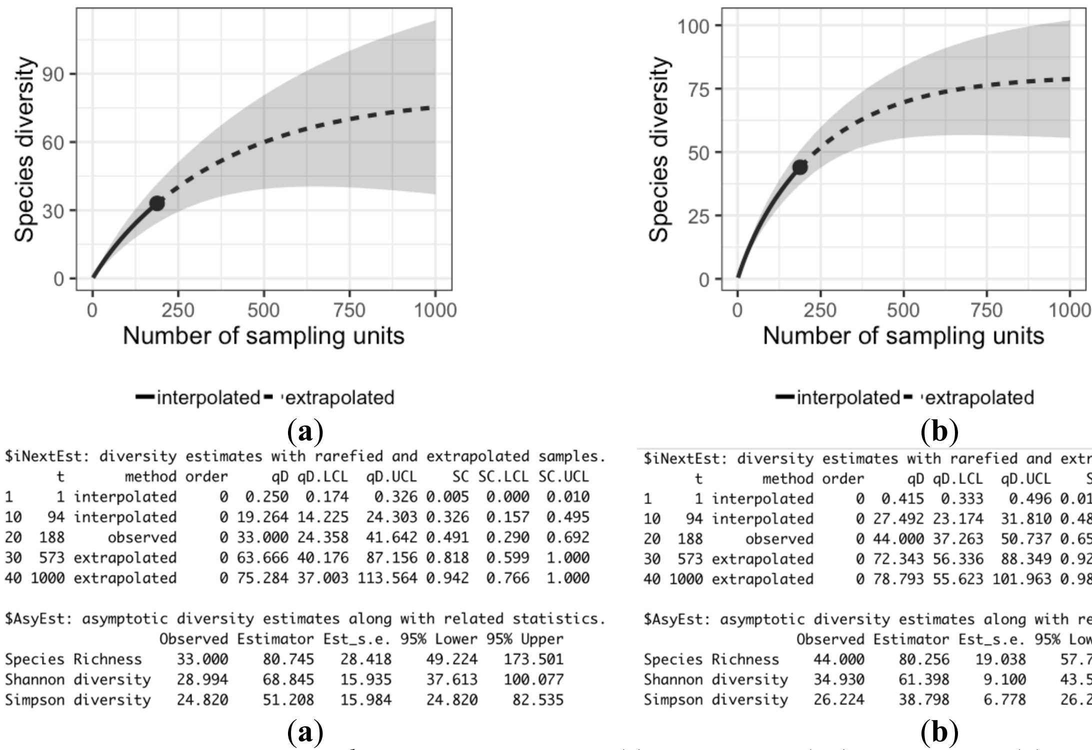
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1 cory	100	75	69	79	81	83	81	81	77	58	57	58	59	58	58	55	53	0	0	0	0	0
2 antinowhere	75	100	68	74	75	79	77	78	72	62	61	61	60	59	61	60	59	0	0	0	0	0
3 crassa	69	68	100	72	73	74	73	73	68	62	61	64	65	62	64	64	62	0	0	0	0	0
4 elmo	80	74	73	100	86	85	87	84	81	53	51	51	51	52	52	50	49	0	0	0	0	0
5 steven	83	76	73	86	100	86	85	86	80	50	50	49	50	49	51	49	46	0	0	0	0	0
6 chumba	83	80	75	84	85	100	89	88	83	50	49	48	49	47	49	47	44	0	0	0	0	0
7 misfit	82	76	74	86	85	89	100	92	82	49	48	46	47	46	49	47	45	0	0	0	0	0
8 billy	81	78	74	84	85	89	92	100	85	48	47	46	46	45	47	47	44	0	0	0	0	0
9 datas	84	79	74	88	87	90	89	92	100	48	48	49	48	48	49	47	44	0	0	0	0	0
10 jett	58	63	63	52	51	50	49	48	45	100	87	83	84	81	85	84	83	0	0	0	0	0
11 sortsol	57	61	61	50	49	48	47	46	44	87	100	83	83	80	83	83	83	0	0	0	0	0
12 willy	59	62	65	50	49	49	47	46	45	84	83	100	94	87	86	86	85	0	0	0	0	0
13 shane	59	61	66	50	50	48	47	46	43	83	83	93	100	88	87	87	86	0	0	0	0	0
14 goonie	59	60	63	52	49	48	47	46	45	82	82	88	90	100	85	84	83	0	0	0	0	0
15 goodold	58	61	65	51	50	50	50	47	45	85	84	85	87	84	100	88	87	0	0	0	0	0
16 debie	56	60	64	50	48	47	48	46	43	84	83	85	87	83	88	100	90	0	0	0	0	0
17 chunk	53	59	63	48	46	45	45	43	40	84	83	84	86	82	87	90	100	0	0	0	0	0
18 fnug	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0
19 iggy	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0
20 oldone	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	23	23
21 clash	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	100	100
22 otherone	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	100	100

Figure S2 phylogenomic nucleotide distances of the 39 unique *P. aeruginosa* phages (Gegenees, BLASTn: fragment size: 200, step size: 100, threshold: 0).

	1	2	3	4	1	2	3	4
1 heks	100	85	0	0	100	90	16	16
2 Nonaheksakonda	80	100	0	0	86	100	16	16
3 vipetofem	0	0	100	64	15	15	100	84
4 nattely	0	0	64	100	15	15	84	100
	(a)				(b)			

Figure S3 (**a**) Phylogenomic nucleotide distances of the 4 unique *Enterococcus* phages phages (Gegenees, BLASTn: fragment size: 200, step size: 100, threshold: 0%). (**b**) Phylogenomic amino acid distances of the 4 unique *Enterococcus* phages (Gegenees, BLASTx: fragment size: 200, step size: 100, threshold: 0%).

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



(a)

$iNextEst: diversity estimates with rarefied and extrapolated samples.

	t	method	order	qD	qD.LCL	qD.UCL	SC	SC.LCL	SC.UCL
1	1	interpolated	0	0.250	0.174	0.326	0.005	0.000	0.010
10	94	interpolated	0	19.264	14.225	24.303	0.326	0.157	0.495
20	188	observed	0	33.000	24.358	41.642	0.491	0.290	0.692
30	573	extrapolated	0	63.666	40.176	87.156	0.818	0.599	1.000
40	1000	extrapolated	0	75.284	37.003	113.564	0.942	0.766	1.000

$AsyEst: asymptotic diversity estimates along with related statistics.

	Observed	Estimator	Est_s.e.	95% Lower	95% Upper
Species Richness	33.000	80.745	28.418	49.224	173.501
Shannon diversity	28.994	68.845	15.935	37.613	100.077
Simpson diversity	24.820	51.208	15.984	24.820	82.535

(a)

(b)

$iNextEst: diversity estimates with rarefied and extrapolated samples.

	t	method	order	qD	qD.LCL	qD.UCL	SC	SC.LCL	SC.UCL
1	1	interpolated	0	0.415	0.333	0.496	0.011	0.004	0.017
10	94	interpolated	0	27.492	23.174	31.810	0.481	0.380	0.582
20	188	observed	0	44.000	37.263	50.737	0.655	0.558	0.752
30	573	extrapolated	0	72.343	56.336	88.349	0.925	0.848	1.000
40	1000	extrapolated	0	78.793	55.623	101.963	0.986	0.955	1.000

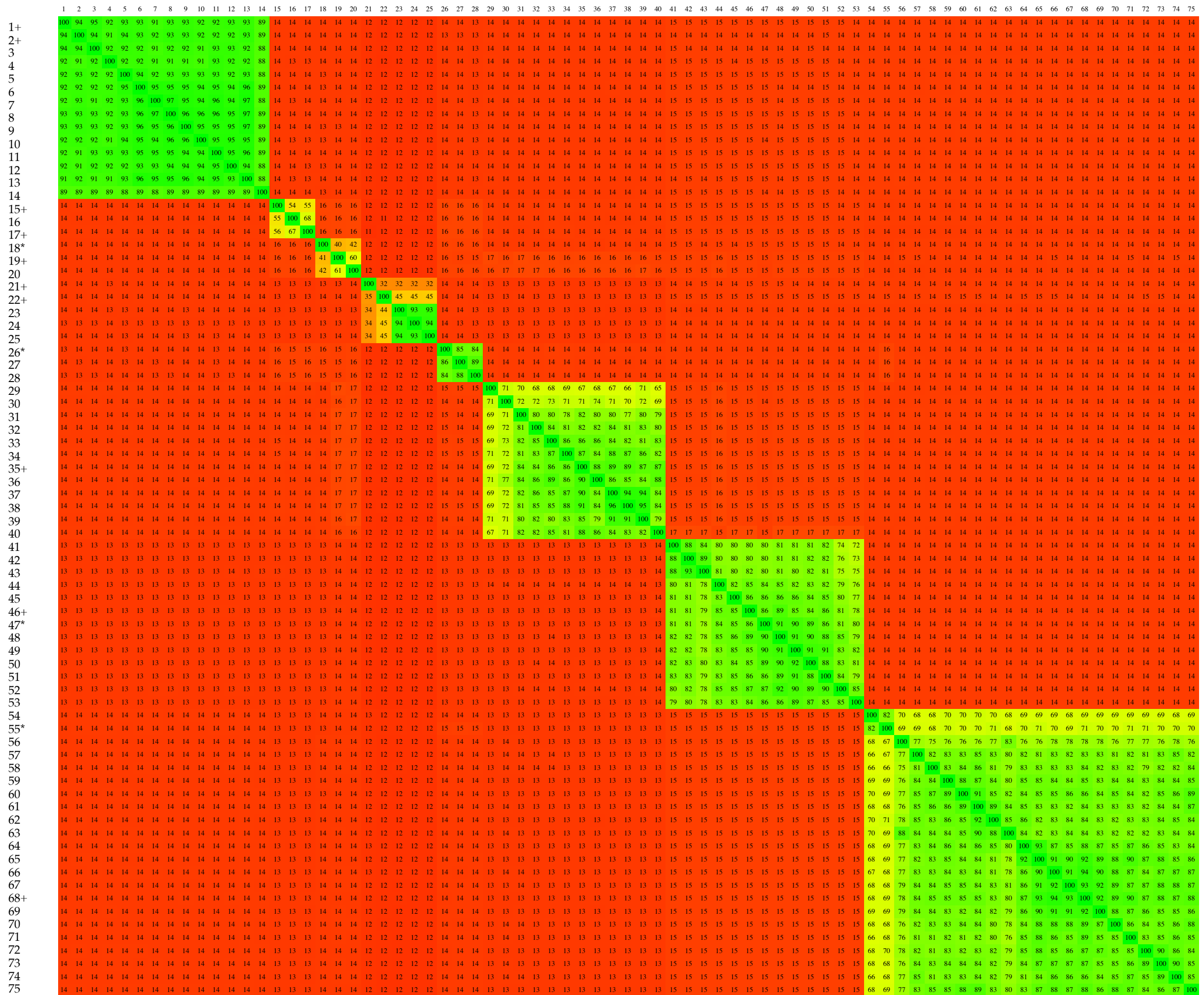
$AsyEst: asymptotic diversity estimates along with related statistics.

	Observed	Estimator	Est_s.e.	95% Lower	95% Upper
Species Richness	44.000	80.256	19.038	57.783	139.374
Shannon diversity	34.930	61.398	9.100	43.563	79.233
Simpson diversity	26.224	38.798	6.778	26.224	52.082

(b)

Figure S4 Rarefaction curves for the *S. enterica* screenings; (**a**) small volume (SV) screening and (**b**) large volume (LV) screening. (**c**) Diversity indices for the small volume (0.5ml) screening based on identification of phages species in all 188 wastewater samples. (**d**) Diversity indices for the large volume (0.5ml) screening based on identification of phages species in all 188 wastewater samples.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.27.011080; this version posted March 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



#	Phage	#	Phage	#	Phage	#	Phage	#	Phage
1	+ Salmonella phage BP63	16	Lumpael	31	dunkel	46	+ Salmonella phage Sh19	61	oselot
2	+ Salmonella phage vB_SenM_PA13076	17	+ Enterobacteria phage IME_EC2	32	celemicas	47	* Escherichia virus CBA120	62	bombadil
3	brunost	18	* Aeromonas phage pIS4-A	33	skrot	48	moki	63	phagemcphageface
4	birk	19	+ Salmonella phage KFS_SE2	34	horsemountain	49	kage	64	misterkot
5	ciri	20	Akira	35	+ Salmonella phage Ent11	50	aagejoakim	65	rokbiter
6	zoltan	21	+ Podoviridae sp. ctdc61	36	wast	51	dinky	66	gmork
7	emiel	22	+ Pectobacterium phage DU_PP_III	37	sidste	52	heyday	67	vaffelhjerte
8	yarpen	23	astrid	38	templet	53	rabagast	68	+ Salmonella phage S147
9	nenneke	24	assan	39	demigod	54	oldekolle	69	polluks
10	rivia	25	astrithr	40	blauehaus	55	* Escherichia virus T5	70	bux
11	emhyr	26	* Shigella phage pSf-1	41	pertopsoe	56	smaug	71	rutana
12	renfre	27	slyngel	42	maane	57	bobsandoy	72	ende
13	triss	28	+ Escherichia phage vB_EcoS_G29-2	43	allotria	58	falkor	73	bastian
14	brorfarstad	29	* Salmonella phage Jersey	44	barely	59	fuchur	74	atrejo
15	+ Escherichia phage Sortsne	30	pink	45	bering	60	faergetype	75	beppo

Figure S5 Phylogenomic amino acid distances of the 59 unique *Salmonella* phages, selected closest relatives and type species of respective genera (Gegenees, BLASTx: fragment size: 200, step size: 100, threshold: 0%). Type species are denoted by an asterisk (*), close relative by plus (+).