Environmental surveillance of viruses by tangential flow filtration and metagenomic reconstruction

V Furtak¹, M Roivainen², O Mirochnichenko¹, T Zagorodnyaya¹, M Laassri¹, SZ Zaidi³, L Rehman³, MM Alam³, V Chizhikov¹, K Chumakov¹

1. Center for Biologics Evaluation and Research, US Food and Drug Administration, Silver Spring, United States

2. National Institute for Health and Welfare, Helsinki, Finland

3. WHO Regional Reference Laboratory for Polio Eradication Initiative, Department of Virology and Immunology, National Institute of Health, Islamabad, Pakistan

Correspondence: Konstantin Chumakov (konstantin.chumakov@fda.hhs.gov)

Citation style for this article:

Furtak V, Roivainen M, Mirochnichenko O, Zagorodnyaya T, Laassri M, Zaidi SZ, Rehman L, Alam MM, Chizhikov V, Chumakov K. Environmental surveillance of viruses by tangential flow filtration and metagenomic reconstruction. *Euro Surveill.* 2016;21(15):pii=30193. DOI: <http://dx.doi.org/10.2807/1560-7917.ES.2016.21.15.30193>

Article submitted on 14 July 2015 / accepted on 14 April 2016 / published on 14 April 2016

An approach is proposed for environmental surveillance of poliovirus by concentrating sewage samples with tangential flow filtration (TFF) followed by deep sequencing of viral RNA. Subsequent to testing the method with samples from Finland, samples from Pakistan, a country endemic for poliovirus, were investigated. Genomic sequencing was either performed directly, for unbiased identification of viruses regardless of their ability to grow in cell cultures, or after virus enrichment by cell culture or immunoprecipitation. Bioinformatics enabled separation and determination of individual consensus sequences. Overall, deep sequencing of the entire viral population identified polioviruses, non-polio enteroviruses, and other viruses. In Pakistani sewage samples, adeno-associated virus, unable to replicate autonomously in cell cultures, was the most abundant human virus. The presence of recombinants of wild polioviruses of serotype 1 (WPV1) was also inferred, whereby currently circulating WPV1 of south-Asian (SOAS) lineage comprised two sub-lineages depending on their non-capsid region origin. Complete genome analyses additionally identified point mutants and intertypic recombinants between attenuated Sabin strains in the Pakistani samples, and in one Finnish sample. The approach could allow rapid environmental surveillance of viruses causing human infections. It creates a permanent digital repository of the entire virome potentially useful for retrospective screening of future discovered viruses.

Introduction

Surveillance for the presence of pathogenic viruses is important for development of rational disease control strategies. It is an integral part of the worldwide campaign to stop the transmission of wild and vaccine-derived polioviruses (VDPV). The campaign uses oral polio vaccine (OPV) made from attenuated strains

of three serotypes of poliovirus that can mutate and become virulent VDPVs. The progress of the campaign is monitored by detecting cases of acute flaccid paralysis (AFP) followed by virological confirmation by serological methods, polymerase chain reaction (PCR) tests, and partial nucleotide (nt) sequencing [1-3]. AFP has infectious and non-infectious causes and therefore polio surveillance involves processing of a large number of samples, most of which do not contain poliovirus. An extensive laboratory network is involved in this activity, making it labour-intensive, time-consuming, and expensive.

The sensitivity of AFP surveillance is inherently limited because in a fully susceptible population only a small fraction of infections leads to AFP. In immunised communities, infection to paralysis ratio is even higher, allowing polioviruses causing AFP to transmit silently [4,5]. Confirmation that poliovirus circulation has stopped must be based on detection of poliovirus itself rather than its clinical manifestations. Poliovirus could be consistently isolated from the environment in regions with no recorded cases of paralytic polio [2,4,6]. Highly evolved VDPVs that may have been excreted by unidentified chronically infected individuals have also been found [7-9].

Current sewage concentration methods are based on two-phase separation [10-12], absorption onto membrane filters [13], and other protocols [14]. While some of these allow only a qualitative characterisation of viruses present in a sample, others, such as two-phase separation can also provide quantitative data. The performance of these methods is however affected by sample pH and the presence of different impurities [15,16]. Traditional size-selective ultrafiltration is impractical because filters are rapidly blocked by impurities. In contrast, devices in which liquid moves tangentially

FIGURE 1

Depth of sequencing coverage and SimPlot analysis of recovered sequences generated by analysis of whole RNA libraries derived from two cell cultures, each infected with a different sewage sample collected in the Helsinki region, Finland, 2013

A. Depth of sequencing coverage generated by analysis of whole RNA library derived from a RD cell culture infected with sewage sample F1. This sewage sample was a priori shown not to contain poliovirus by a standard World Health Organization procedure using strain-specific polymerase chain reaction [21]. Red and blue indicate forward and reverse sequencing reads, respectively.

B. Depth of sequencing coverage generated by analysis of whole RNA library derived from a L2oB cell culture infected with sewage sample F2. This sewage sample was a priori shown to contain poliovirus. Red and blue indicate forward and reverse sequencing reads, respectively.

C. SimPlot diagram generated using consensus sequences of the virus present in F1 sewage sample compared with coxsackievirus B3 and echovirus 11, accession numbers AY752944 and AJ276224.

D. SimPlot diagram generated using consensus sequences of the virus present in F2 sewage sample compared with Sabin 2 and Sabin 3, accession numbers AY184220 and AY184221.

to the filter surface (tangential flow filtration, TFF) prevent the significant reduction of flow rates. TFF was successfully used for concentration of viruses, bacteria, and parasites from environmental samples [17-19].

Conventional protocols for the identification of polioviruses in sewage concentrates include cultivation in cell cultures susceptible to poliovirus [12,20], followed by identification using strain-specific PCR and partial sequencing of their genomes [21]. The process is quite laborious and is often complicated by the presence of a mixture of viruses. Massively parallel or deep sequencing generates vast amounts of sequence information and is often used for the analysis of heterogeneous viral populations and metagenomic studies. Metagenomics

have many applications in the environmental sciences [22-24] and were also successfully used for the identification of respiratory viruses in clinical samples [25]. Here we explore the utility deep sequencing for the analysis of TFF-concentrated environmental samples.

Methods

Tangential flow ultrafiltration of sewage samples

0.5 L sewage samples were collected in 2013 in the Helsinki region of Finland (two samples) and Islamabad, Pakistan (eight samples). They were centrifuged for 10 min at 3,000 *g* resulting in a solid and liquid fraction. The solid (sludge) fraction was

FIGURE 2

Comparison of sequence diversity profiles in two sewage samples with heterogeneous and homogeneous viral populations, Finland 2013

A. Sequence profiles of the F2 sewage sample containing a virus ca 16% distant from the major viral population.

B. Sequence profile of another sewage sample with a more homogenous viral population

re-suspended and homogenised by Omni homogenizer in 25 mL of 25% beef extract, and the resulting particulate fraction removed by centrifugation at 3,000 *g* for 10 min. The supernatant was combined with the previously obtained liquid fraction and filtered through a 0.2 µm MidiKros TFF filter (Spectrum Laboratories Inc., Rancho Dominguez, CA) followed by rinsing the filter with 10 mL of 25% beef extract. In some cases liquid and solid (i.e. supernatant obtained from elution with beef extract) fractions were processed separately. Next, virus concentration was performed using an automated Spectrum Laboratories KrosFlo Research Ili TFF system and 100 kDa molecular weight cut-off MidiKros filters at 9 psi trans-membrane pressure and 60 mL/min flow rate to the final volume of ca 5 mL (the void volume of the TFF system). The volume was further reduced to 0.5–0.8 mL by manually operating two 5-mL syringes attached to TFF cartridge disconnected from the TFF system.

Isolation of viruses in cell cultures

Sewage concentrates were diluted in Dulbecco minimal essential medium (DMEM) to a final volume of 3 mL and inoculated into cultures of rhabdomyosarcoma (RD) cells and incubated in a 5% CO₂ atmosphere at 37°C for 1–7 days until cytopathic effect (CPE) developed. The harvest from RD cultures was inoculated into L20B cell cultures (mouse L cells expressing human poliovirus receptor CD155) to selectively amplify polioviruses according to the World Health Organization (WHO) protocol [13]. Culture media from CPE-positive L20B cells and CPE-positive RD cells were used for subsequent real-time PCR analysis and/or deep sequencing.

Detection of polio and enteroviruses in sewage concentrates by real-time PCR

RNA was isolated from samples prepared by different procedures described above using NucleoSpin RNA Virus F kit (Clontech Inc., Mountain View, CA) and used for real-time PCR with strain-specific primers [25]. All PCR reactions were performed in duplicate on a CFX96 Real-Time PCR system (Bio-Rad). The same RNA samples were also used for deep sequencing.

Titration of Sabin 1 poliovirus in spiked sewage samples

0.5 mL of 2x10⁵ cell culture infectious doses (CCID₅₀) of Sabin 1 poliovirus diluted in phosphate buffered saline (PBS) was added to 0.5 L of poliovirus-free sewage sample and incubated for 1 hour. Identical 0.5 mL virus sample served as a control. Virus titres were determined in Hep2C cells by terminal dilutions method according to WHO recommendations [13]. Virus recovery was evaluated as a percentage of infectious virus relative to the control unspiked sample.

Poliovirus enrichment by immunoprecipitation and sequence-independent isothermal amplification

Biotinylated rabbit polyclonal IgG against wild polioviruses of serotypes 1, 2, and 3 were used for immunoprecipitation. Five µL of stock solution containing 0.5 µg rabbit IgG was added to 0.5 mL of TFF-concentrated sewage and incubated for 3 hours on a rocker. Ten µL of streptavidin-coated magnetic beads (Thermo Fisher Scientific Inc., Rockford, IL) washed with PBS were added to the TFF concentrate and incubated at room temperature for 2 hours on a rocker. Magnetic beads were then washed with DMEM and PBS and

FIGURE 3

Phylogenetic tree showing the relatedness of viral sequences identified in F1 and F2 sewage samples to reference enterovirus strain sequences, Helsinki region, Finland, 2013

Sequences marked sp1 and sp2 were bioinformatically separated from sewage sample F2, and represent the majority and minority component, respectively.

FIGURE 4

Simplot analyses showing recombinant structures of some polioviruses present in sewage samples collected in Islamabad, Pakistan, June–July 2013

Reference sequences used for these Simplot charts were Sabin 1 (GenBank accession number: AY184219), Sabin 2 (GenBank accession number: AY184220), Sabin 3 (GenBank accession number: AY184221), and EGY1218588 (GenBank accession number: KJ155496).

A. Recombinant Sabin 1/2 virus.

B. Wild type 1 poliovirus, which, when compared to sequence entries in the entire database of published enterovirus sequences, was most similar to EGY1218588 (which belongs to the south-Asian (SOAS) lineage).

C. Recombinant Sabin 3/2 virus.

D. Wild type 1 poliovirus, which, when compared to sequence entries in the entire database of published enterovirus sequences, was most similar to EGY1218588 (which belongs to the SOAS lineage). The similarity between the virus in panel D and EGY1218588 was less than the similarity between the virus in panel B and EGY1218588. The difference between the two wild type viruses in panels B and D could be explained by a recombination event in the sequence of the virus in panel D at the boundary between the P1 and P2 regions.

stored frozen at -20°C. The procedure was performed separately for each serotype of poliovirus. Isothermal amplification using REPLI-g WTA single cell kit (Qiagen, Valencia, CA) was performed according to the manufacturer's protocol.

Deep sequencing

RNA isolated from samples prepared by different procedures described above was used for total RNA Illumina library preparation using NEB NEXT Ultra RNA library kit (NEB, Ipswich, MA). The quality of libraries was assessed using the qPCR Library Quantification Kit (Kapa Biosystems, Inc., Woburn, MA) and Agilent 2100

Bioanalyzer (Agilent Technologies, Santa Clara, CA). DNA sequencing was performed on Illumina HiSeq 2000 or MiSeq to generate 100 to 300 nt long paired-end reads. In some experiments, samples were sequenced in multiplex format (up to eight samples per sequencing lane) using unique index primers (barcodes).

Bioinformatic analysis

Bioinformatic analysis was performed using a custom 'swarm' software package developed for UNIX (Darwin) environment on MacPro (Apple Inc., Cupertino, CA) and running in parallel mode using up to 22 computation threads. Low quality sequences, adapters and index

TABLE 1

A list of viruses identified in eight sewage samples collected in Islamabad, Pakistan, June–July 2013

Sewage sample name	Fraction	Cell culture	Majority virus	Minority virus ^a	
				% ^b	Description
P1	Supernatant	RD	Wt polio type 1	15	Sabin 3
		L20B	Wt polio type 1	5	Sabin 3
	Sludge	RD	Coxsackievirus B5	50	Sabin 1
		L20B	Wt polio type 1	5	Wt Polio type 1
	Combined	RD	Wt polio type 1	40	Sabin 1, Sabin 3
		L20B	Wt polio type 1	2	Sabin 1
P2	Supernatant	RD	Wt polio type 1	20	Sabin 3
		L20B	Wt polio type 1	15	Sabin 3
	Sludge	RD	Wt polio type 1	30	Coxsackievirus B1
		L20B	Wt polio type 1	40	Wt Polio type 1
	Combined	RD	Wt polio type 1	35	Sabin 1
		L20B	Wt polio type 1	15	Sabin 3, Sabin 1
P3	Combined	RD	Coxsackievirus B5	10	Sabin 3, Sabin 1
		L20B	Sabin 2	25	Coxsackievirus B1, B3, B5
P4	Combined	RD	Sabin 2	30	Coxsackievirus B3
		L20B	Sabin 3	20	Coxsackievirus B5
P9	Supernatant	RD	Sabin 3	15	Coxsackievirus B5
		L20B	Sabin 3	— ^c	— ^c
	Sludge	RD	Coxsackievirus B5	20	Sabin 3
		L20B	Sabin 3	40	Wt Polio type 1
	Combined	RD	Coxsackievirus B5	50	Sabin 3, Sabin 3/1 recombinant
		L20B	Sabin 3	10	Coxsackievirus B5, Sabin 3/1 recombinant
P10	Supernatant	RD	Coxsackievirus B5	30	Coxsackievirus B5
		L20B	Wt polio type 1	— ^c	— ^c
	Sludge	RD	Sabin 2	15	Coxsackievirus B3, B5
		L20B	Sabin 2	5	Sabin 1
	Combined	RD	Sabin 2	50	Coxsackievirus B5
		L20B	Sabin 2	35	Coxsackievirus B5
P11	Supernatant	RD	No product	— ^c	— ^c
		L20B	Wt polio type 1	— ^c	— ^c
	Sludge	RD	No product	— ^c	— ^c
		L20B	Wt polio type 1	— ^c	— ^c
	Combined	RD	Wt polio type 1	— ^c	— ^c
		L20B	Wt polio type 1	45	Wt Polio type 1
P12	Supernatant	RD	Wt polio type 1	50	Sabin 2
		L20B	Sabin 2	— ^c	— ^c
	Sludge	RD	No product	— ^c	— ^c
		L20B	Sabin 1	20	Sabin 3
	Combined	RD	Sabin 1	50	Sabin 1
		L20B	Sabin 1	25	Sabin 3

Wt: wild type.

^a The minority virus is inferred from the finding, next to the majority consensus sequence (of the 'majority' virus), of another consensus sequence which is reconstructed after separation of deep sequence reads using the algorithm described in the text.^b Percentage indicates the proportion of deep sequence reads that match to the alternative consensus(es).^c Only one virus (i.e. the 'majority' virus) was found in the sewage fraction considered, so there is no 'minority' virus for this fraction.

TABLE 2

Relative abundance of the most prevalent RNA viruses present in a sewage sample collected in Pakistan, 2013

Virus	Proportions of reads among all reads that match the sequence of a given virus (%)
Tomato mosaic virus	41.20
Cucumber green mottle mosaic virus	26.10
Tobacco mild green mosaic virus	14.20
Pepper mild mottle virus	6.70
Tobacco mosaic virus	6.00
Youcai mosaic virus	2.00
Tomato mottle mosaic virus isolate MX5	1.10
Adeno-associated virus - 2	0.50
Melon necrotic spot virus	0.30
Poliovirus	0.20
Other	1.50

sequences were removed, and the remaining reads were aligned (mapped) on curated databases of full-length sequences of different virus groups (GenBank). The mapping procedure involved k-mer guided Smith–Waterman alignment that will be described elsewhere. Profiles of sequence heterogeneity were computed, and consensus sequences were built by identifying the most frequent nts at each genomic position. In cases where significant sequence heterogeneity was detected, sequence reads were separated into discrete sub-populations using the following algorithm. Sequence reads mapped against aligned reference genomes were segregated into discrete subsets by cluster analysis performed in overlapping ‘windows’ of 50–100 nt covering the entire genome (‘sliding window’ method). The identity of each species defined in a particular window was matched with similar information obtained for the adjacent overlapping window, and propagated along the entire genome. Consensus sequences were computed for each subset of sequence reads as described above. Unique sequences of wild type polioviruses determined in this study were deposited in GenBank (accession numbers: KU161395–KU161399).

Results

Concentration of sewage samples by tangential flow filtration and recovery of viruses

The entire concentration procedure of 0.5 L sewage samples took ca 2 hours. Final sewage concentrates in 0.5–0.8 mL volume were used for RNA isolation (see above) or inoculation into cell cultures.

Analysis by the standard WHO procedure using strain-specific PCR [21] had previously shown that a sewage sample hereby named ‘F1’ and collected in Finland did

not contain poliovirus, while another sample ‘F2’, from the same country, was positive for Sabin 2 poliovirus.

To determine the efficiency of the TFF protocol, the F1 sample was spiked with 10^5 CCID₅₀ of Sabin 1 poliovirus. The recovery of spiked poliovirus in the combined sludge/supernatant fraction was 62%, 84%, and 93% in three independent experiments.

Sensitivity determined in spiking experiments may differ from the sensitivity of virus detection in real sewage samples. Therefore to validate the concentration method, another sewage sample F2 from Finland which was demonstrated to contain Sabin 2 poliovirus was processed by the TFF protocol. The concentrates of F1 (poliovirus-free) and F2 were used to infect RD and then L20B cell cultures. RD cultures were virus positive for both samples, while only F2 induced CPE in L20B cells.

Deep sequencing of cell culture-grown viruses

Whole-RNA Illumina libraries prepared from supernatants of RD and L20B cultures infected with F1 and F2 concentrates were deep-sequenced using Illumina HiSeq2000. Individual sequence reads were aligned to genomic sequences of 744 representative enteroviruses. Figures 1A and 1B show the depth of sequencing coverage along the viral genome in forward and reverse orientations. The most prevalent nt at each genomic position was considered to represent the consensus, and all others were assumed to be single nt polymorphisms (SNP). Consensuses contained uninterrupted open reading frames coding for 2,196 and 2,208 amino acids for F1 and F2 samples.

Their comparison with references and SimPlot analysis revealed that the closest relative of the virus present in the F1 sample was echovirus 11 with 82% similarity in the P1 region (Figure 1C). This level of homology does not allow us to unambiguously identify the serotype of this virus.

The virus from the F2 sample was a recombinant of Sabin 2 and Sabin 3 viruses with the crossover at the boundary between regions coding for 3C protease and 3D polymerase (nt 5,959–5,981; Figure 1D). The distal part of the genome was identical to the Sabin 3 virus, while the proximal part differed from the Sabin 2 virus at six nt positions (two in 5'-untranslated region (UTR), two in viral protein VP2, one in VP1, and one in 2A protein). This indicates that this was a minimally evolved vaccine poliovirus, however, it contained mutations at both sites associated with neurovirulence ($A_{481} \rightarrow G$ in 5'-UTR and $Thr_{143} \rightarrow His$ in VP1).

Sequences determined by the above procedure represent consensuses of the majority component. Sequence heterogeneity profiles revealed that sewage sample F2 contained ca 16% of one or more viruses distinct from the predominant Sabin 2/3 recombinant (Figure 2A). For comparison, Figure 2B shows an example of the

TABLE 3

Relative proportion of sequence reads matching the most abundant viruses detected in a sewage sample before and after enrichment by immunoprecipitation using antibodies against poliovirus

Virus found before enrichment	Per cent ^a	Virus found after enrichment	Per cent ^a
Tomato mosaic virus	41.20	Poliovirus	8.90
Cucumber green mottle mosaic virus	26.10	Tobacco mild green mosaic virus	8.30
Tobacco mild green mosaic virus	14.20	Circovirus-like genome RW-C	6.70
Pepper mild mottle virus	6.70	Cucumber green mottle mosaic virus	4.30
Tobacco mosaic virus	6.00	Tobacco mosaic virus	2.70
Youcai mosaic virus	2.00	Tomato mottle mosaic virus isolate MX5	2.70
Tomato mottle mosaic virus isolate MX5	1.10	Lactococcus phage phiLC3	2.00
Adeno-associated virus - 2	0.50	Sulfolobus islandicus rod-shaped virus 1	2.00
Melon necrotic spot virus	0.30	Zamilon virophage	1.90
Poliovirus	0.20	Enterococcus phage EF62phi	1.70
Cosavirus A strain HCoV-A1	0.10	Cotesia congregata virus	1.70
Bovine viral diarrhoea virus 1	0.10	Clostridium phage phiMMPo4	1.50
Other	0.80	Other	54.50

^a The percentages represent the proportions of reads among all reads that match the sequence of a given virus.

SNP profile of a relatively homogeneous viral sample. To separate viruses present in the mixture, individual sequence reads mapped on the aligned reference dataset were subjected to cluster analysis in a 'sliding window' format (in a series of overlapping 50–100 nt-long sections). Sequence reads separated by this procedure were assembled into full-genome consensus sequences. The majority species in F2 samples grown in L20B cells was identical to Sabin 2/3 recombinant determined by building the overall consensus. The minority species in F2 sample was close to echovirus 11 virus with ca 10% differences from the virus identified in F1 sample (Figure 3).

Analysis of sewage samples from polio-endemic environment

The above results established the feasibility of using TFF concentration and deep sequencing for characterising multiple viruses present in viral populations. Next, the utility of this approach for surveillance in a country that is still endemic for poliovirus was explored. Eight sewage samples collected in Pakistan were processed as described above, concentrates were inoculated into RD and L20B cultures, and harvested viruses were deep-sequenced. Multiple enteroviruses present in most samples were separated bioinformatically as described above. Table 1 lists discrete viruses found in these samples. We found viruses that were very close to all three Sabin vaccine strains, as well as a number of wild type 1 polioviruses. In some samples, there were also sequences related to non-polio enteroviruses, predominantly of species B (coxsackievirus B1, B3, and B5).

The approach used in this work to reconstruct genomic sequences resulted in complete or near complete

genomes (in some cases missing 1–10 nt at the ends, due to low sequencing coverage of the terminal regions). It produced more phylogenetic and molecular-epidemiological information than would be available based on VP1 sequencing alone. For instance, some vaccine derivatives were found to be recombinants of two serotypes of Sabin virus, and the crossover point was easily identified by SimPlot analysis (Figure 4A,4C). Similarly, wild type 1 of south-Asian (SOAS) lineage currently prevalent in Pakistan and Afghanistan contained two sub-lineages that could be distinguished by a recombination event at the boundary between the P1 and P2 regions (Figure 4B,4D).

Metagenomic approach

TFF concentration followed by deep sequencing of viruses isolated in cell culture represents an improvement over the conventional protocol by providing significantly more information in a much shorter time. However, it still depends on the ability of viruses to grow in cell culture, potentially biasing the result. A totally agnostic approach based on the direct sequencing of all RNA present in sewage concentrates could produce a more complete picture. Whole-RNA Illumina libraries were prepared directly from TFF concentrates of sewage collected in Pakistan and deep sequenced. Sequence reads were aligned to reference sequences of 5,301 full length viral genomes belonging to all major groups of viruses (National Center for Biotechnology Information, Bethesda, MD). Approximately 5–8% of all sequence reads could be aligned to sequences in the viral database, and most hits belonged to plant viruses (Table 2). Separation of sequences of plant viruses into discrete subpopulations by the algorithm described above revealed the presence of full length genomes of several different viruses, including tomato mosaic

virus, cucumber green mottle mosaic virus, tobacco mild green mosaic virus, pepper mild mottle virus, tobacco mosaic virus, Youcai mosaic virus, and tomato mottle virus (Table 2).

Surprisingly, the most represented of all animal viruses was adeno-associated virus (AAV, 0.5% of all reads) followed by poliovirus (0.2%). Mapping of sequence reads by alignment to sequences in reference databases of individual virus families that produced a substantial number of hits during the preliminary screening, allowed the assembly of full-genome sequences of viruses present in the sample. Sequence reads aligned to sequences in the database containing 18 reference AAV sequences assembled into a full-length genome of 4,675 nt containing two open reading frames corresponding to the non-structural and the capsid proteins of the virus. Phylogenetic analysis showed that it was 97.3% similar to AAV-2. A similar analysis performed for astroviruses revealed the presence of human astrovirus 4, Aichi virus 1, human parechovirus 1, as well as several other mammalian RNA and DNA viruses (data not shown).

Enrichment of tangential flow filtration-concentrated sewage by immunoprecipitation

The number of hits for most other groups of viruses was insufficient to generate good coverage allowing unambiguous assembly of complete viral genomes or reliable heterogeneity analysis. To enable a more sensitive search for specific viruses, polioviruses present in TFF concentrates were enriched by immunoprecipitation. TFF-concentrated sewage samples were incubated with biotinylated rabbit polyclonal anti-polio IgG and the virus was captured using streptavidin-coated microbeads. Complementary DNA (cDNA) from these samples was isothermally amplified using a uniform sequence-independent REPLI-g system. Table 3 shows the specificity of sequence reads from one sewage sample before and after the enrichment that resulted in poliovirus sequences becoming the predominant species. The consensus sequence assembled from these sequence reads was shown by phylogenetic analysis to be of SOAS WPV1 identical to the virus isolated from the same sewage sample in cell culture.

Discussion

This communication demonstrates that TFF can effectively concentrate viruses present in sewage by reducing the volume from ca 500 mL to ca 1 mL in approximately 2 hours. Unlike the conventional protocols [10,12,21] it does not require large volumes of toxic and noxious organic solvents, removes most impurities including PCR inhibitors, and is more amenable to high-throughput implementation.

Traditionally, identification of viruses in sewage concentrates is performed by using selective cell cultures such as mouse L20B cells that express human poliovirus receptor [20], followed by strain-specific PCR and partial nt sequencing [10,12,21]. The improvements

proposed here include (i) the use of deep sequencing for virus identification, and (ii) direct metagenomic analysis of sewage concentrates allowing viruses to be identified without cell cultures. The main advantage of deep sequencing is that it is a universal procedure that detects multiple viruses present in mixtures, and generates complete or near-complete genome sequences that contain more molecular-epidemiological information than partial sequencing.

Thus several vaccine-related and wild type polioviruses identified in this study were found to be intertypic recombinants. Finding the recombinant Sabin 2/3 poliovirus in sewage from Finland was unexpected because this country does not use oral polio vaccine (OPV). Even though this 'young' vaccine-derivative contained only six mutations compared with Sabin 2, two of them were previously linked to neurovirulent reversion. Whole genome analysis of wild polioviruses revealed that currently circulating strains in Pakistan belong to at least two sub-lineages distinguished by different origins of their P2-P3 region (Figure 4B,4D).

Most samples contained more than one virus, including non-polio enteroviruses, predominantly of species B, similar to coxsackievirus B1, B3, B5, and some echoviruses. However, since no sufficiently close match was found in the published database, their serotype identity could not be established unambiguously. Future expansion of the sequence database could enable more accurate identification. Nevertheless, our study demonstrates that this approach could be also used for the surveillance of all non-polio enteroviruses.

Deep-sequencing of the entire RNA content of sewage concentrates also revealed the prevalence of multiple plant viruses, consistent with previous studies of human stool samples [26]. Bioinformatic separation of deep-sequence reads belonging to different viruses allowed their assembly into several full-length genomes that were then identified by phylogenetic analysis. Surprisingly, the most prevalent animal virus in sewage from Pakistan was AAV-2. It has never been associated with any known pathology, and cannot grow in cell cultures without a helper adenovirus [27]. The epidemiology of this virus is unknown, and this observation could help to identify its niche in nature. Additional animal viruses were detected in uncultivated sewage concentrates collected in Pakistan, and their complete genomes sequenced. They included wild and attenuated polioviruses and other enteroviruses described in this paper, as well as cosaviruses, astroviruses, parechoviruses, kobuvirus (Aichi), Saffold virus, bovine viral diarrhoea virus, cricket paralysis virus, and others that were not included in this communication and will be described elsewhere.

The most intriguing aspect of this study is the possibility of a direct metagenomic approach for routine surveillance, bypassing the need for viral cell cultures. The use of this 'agnostic' metagenomic approach

additionally raises an interesting possibility of future retrospective search for viruses unknown at the time of sample collection by aligning sequences determined in the past to reference sequences of newly discovered viruses. Therefore, the deep sequence data could serve as a digital repository of the virome present in sewage samples without the need to preserve their viability or biocontainment.

Depending on the sewage sample, a certain number of viruses were represented by a small number of sequence reads. Although these could have potentially resulted from contamination, such as carryover of a small amount of genetic material from previous sequencing runs in the laboratory, we do not believe that findings described in this report resulted from such carryover. Indeed for the viral sequences that could be reconstructed, we do not work with the corresponding viruses in our laboratory. Moreover, bleaching between sequencing runs, which helps to eliminate this problem, was regularly performed.

Low number of reads of certain viruses, however make the depth of the sequencing coverage insufficient for reliable assembly of their complete genomes. Enrichment of polioviruses from sewage concentrates by immunoprecipitation resulted in poliovirus-specific sequence reads becoming predominant in uncultivated sewage concentrates, preserving the remaining material for further screening for other viruses. As a result, the complete genome of wild type 1 poliovirus was assembled and was found to match the sequence of the virus of the SOAS lineage that was identified in this sample by the cell culture method.

In conclusion, the approach described in this paper can provide an effective alternative to current protocols used for environmental surveillance for wild polioviruses and VDPV, other non-polio enteroviruses, as well as viruses of other families. Most countries in Europe have switched to the exclusive use of Inactivated Polio Vaccines (IPV). Good protection from paralysis that it induces combined with inadequate intestinal immunity [28] allows viruses to spread silently, which makes the traditional AFP surveillance less effective. Therefore environmental surveillance using metagenomic approach in combination with TFF or other improved sewage concentration methods could be particularly useful in Europe and other countries that have switched to the use of IPV. Rapidly declining cost of deep sequencing makes this approach cost-effective compared with the conventional sequencing methods. It could significantly shorten the time needed for the identification of pathogenic viruses in the environment, allowing for a much faster deployment of countermeasures.

Erratum

The name of the author Sohail Z. Zaidi had been misspelled. This was corrected on 12 May 2016.

Acknowledgements

Authors thank Dr. Humayun Asghar of the World Health Organization for suggestions. This work was funded by the intramural research program of the US Food and Drug Administration and a grant to M.R. from the World Health Organization.

Conflict of interest

None declared.

Authors' contributions

Konstantin Chumakov and Merja Roivainen designed the study. Merja Roivainen, Sohail Z. Zaidi, Lubna Rehman, and Muhammad M. Alam collected sewage samples, Vyacheslav Furtak, Merja Roivainen, and Vladimir Chizhikov performed TFF concentration, Olga Mirochnichenko performed virus isolation, Tatiana Zagorodnyaya, Majid Laassri, and Vyacheslav Furtak performed deep sequencing. Konstantin Chumakov drafted the manuscript.

References

- Centers for Disease Control and Prevention (CDC). Evaluating surveillance indicators supporting the Global Polio Eradication Initiative, 2011-2012. *MMWR Morb Mortal Wkly Rep*. 2013;62(14):270-4. PMID: 23575241
- Hovi T, Shulman LM, van der Avoort H, Deshpande J, Roivainen M, DE Gourville EM. Role of environmental poliovirus surveillance in global polio eradication and beyond. *Epidemiol Infect*. 2012;140(1):1-13. DOI: 10.1017/S095026881000316X PMID: 21849095
- Asghar H, Diop OM, Weldegebriel G, Malik F, Shetty S, El Bassioni L, et al. Environmental surveillance for polioviruses in the Global Polio Eradication Initiative. *J Infect Dis*. 2014;210(Suppl 1):S294-303. DOI: 10.1093/infdis/jiu384 PMID: 25316848
- Manor Y, Handscher R, Halmut T, Neuman M, Bobrov A, Rudich H, et al. Detection of poliovirus circulation by environmental surveillance in the absence of clinical cases in Israel and the Palestinian authority. *J Clin Microbiol*. 1999;37(6):1670-5. PMID: 10325305
- Shulman LM, Gavrilin E, Jorba J, Martin J, Burns CC, Manor Y, et al. Molecular epidemiology of silent introduction and sustained transmission of wild poliovirus type 1, Israel, 2013. *Euro Surveill*. 2014;19(7):20709. DOI: 10.2807/1560-7917.ES2014.19.7.20709 PMID: 24576471
- El Bassioni L, Barakat I, Nasr E, de Gourville EM, Hovi T, Blomqvist S, et al. Prolonged detection of indigenous wild polioviruses in sewage from communities in Egypt. *Am J Epidemiol*. 2003;158(8):807-15. DOI: 10.1093/aje/kwg202 PMID: 14561671
- Shulman LM, Manor Y, Sofer D, Handscher R, Swartz T, Delpeyroux F, et al. Neurovirulent vaccine-derived polioviruses in sewage from highly immune populations. *PLoS One*. 2006;1(1):e69. DOI: 10.1371/journal.pone.0000069 PMID: 17183700
- Al-Hello H, Jorba J, Blomqvist S, Raud R, Kew O, Roivainen M. Highly divergent type 2 and 3 vaccine-derived polioviruses isolated from sewage in Tallinn, Estonia. *J Virol*. 2013;87(23):13076-80. DOI: 10.1128/JVI.01174-13 PMID: 24049178
- Hovi T, Paananen A, Blomqvist S, Savolainen-Kopra C, Al-Hello H, Smura T, et al. Characteristics of an environmentally monitored prolonged type 2 vaccine derived poliovirus shedding episode that stopped without intervention. *PLoS One*. 2013;8(7):e66849. DOI: 10.1371/journal.pone.0066849 PMID: 23935826
- Albertson P. Two-phase separation of viruses. In Maramorosch K, Koprowski H, eds. *Methods of virology*. 1967;303-21.
- Más Lago P, Gary HE, Pérez LS, Cáceres V, Olivera JB, Puentes RP, et al. Poliovirus detection in wastewater and stools following an immunization campaign in Havana, Cuba. *Int J Epidemiol*. 2003;32(5):772-7. DOI: 10.1093/ije/dyg185 PMID: 14559748

12. World Health Organization (WHO). Guidelines for environmental surveillance of poliovirus circulation. Geneva, Switzerland: WHO; 2003.
13. Tambini G, Andrus JK, Marques E, Boshell J, Pallansch M, de Quadros CA, et al. Direct detection of wild poliovirus circulation by stool surveys of healthy children and analysis of community wastewater. *J Infect Dis.* 1993;168(6):1510-4. DOI: 10.1093/infdis/168.6.1510 PMID: 8245537
14. Lewis GD, Metcalf TG. Polyethylene glycol precipitation for recovery of pathogenic viruses, including hepatitis A virus and human rotavirus, from oyster, water, and sediment samples. *Appl Environ Microbiol.* 1988;54(8):1983-8. PMID: 2845860
15. Shields PA, Farrah SR. Concentration of viruses in beef extract by flocculation with ammonium sulfate. *Appl Environ Microbiol.* 1986;51(1):211-3. PMID: 3006587
16. Guttman-Bass N, Catalano-Sherman J. Humic acid interference with virus recovery by electropositive microporous filters. *Appl Environ Microbiol.* 1986;52(3):556-61. PMID: 3021058
17. Liu P, Hill VR, Hahn D, Johnson TB, Pan Y, Jothikumar N, et al. Hollow-fiber ultrafiltration for simultaneous recovery of viruses, bacteria and parasites from reclaimed water. *J Microbiol Methods.* 2012;88(1):155-61. DOI: 10.1016/j.mimet.2011.11.007 PMID: 22108496
18. Smith CM, Hill VR. Dead-end hollow-fiber ultrafiltration for recovery of diverse microbes from water. *Appl Environ Microbiol.* 2009;75(16):5284-9. DOI: 10.1128/AEM.00456-09 PMID: 19561183
19. Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. Laboratory procedures to generate viral metagenomes. *Nat Protoc.* 2009;4(4):470-83. DOI: 10.1038/nprot.2009.10 PMID: 19300441
20. Pipkin PA, Wood DJ, Racaniello VR, Minor PD. Characterisation of L cells expressing the human poliovirus receptor for the specific detection of polioviruses in vitro. *J Virol Methods.* 1993;41(3):333-40. DOI: 10.1016/0166-0934(93)90022-J PMID: 8386181
21. Kilpatrick DR, Yang CF, Ching K, Vincent A, Iber J, Campagnoli R, et al. Rapid group-, serotype-, and vaccine strain-specific identification of poliovirus isolates by real-time reverse transcription-PCR using degenerate primers and probes containing deoxyinosine residues. *J Clin Microbiol.* 2009;47(6):1939-41. DOI: 10.1128/JCM.00702-09 PMID: 19386844
22. Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, et al. Functional metagenomic profiling of nine biomes. *Nature.* 2008;452(7187):629-32. DOI: 10.1038/nature06810 PMID: 18337718
23. Cantalupo PG, Calgua B, Zhao G, Hundesa A, Wier AD, Katz JP, et al. Raw sewage harbors diverse viral populations. *MBio.* 2011;2(5):e00180-11. DOI: 10.1128/mBio.00180-11 PMID: 21972239
24. Ng TF, Marine R, Wang C, Simmonds P, Kapusinszky B, Bodhidatta L, et al. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. *J Virol.* 2012;86(22):12161-75. DOI: 10.1128/JVI.00869-12 PMID: 22933275
25. Prachayangprecha S, Schapendonk CM, Koopmans MP, Osterhaus AD, Schürch AC, Pas SD, et al. Exploring the potential of next-generation sequencing in detection of respiratory viruses. *J Clin Microbiol.* 2014;52(10):3722-30. DOI: 10.1128/JCM.01641-14 PMID: 25100822
26. Zhang T, Breitbart M, Lee WH, Run JQ, Wei CL, Soh SW, et al. RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol.* 2006;4(1):e3. DOI: 10.1371/journal.pbio.0040003 PMID: 16336043
27. Mueller C, Flotte TR. Clinical gene therapy using recombinant adeno-associated virus vectors. *Gene Ther.* 2008;15(11):858-63. DOI: 10.1038/gt.2008.68 PMID: 18418415
28. Plotkin SA, Vidor E. Poliovirus vaccine -- inactivated. In: Plotkin SA, Orenstein WA, Offit PA, editors. *Vaccines.* 5 ed. Philadelphia: W.B. Saunders; 2008. p. 605-29.

License and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence, and indicate if changes were made.

This article is copyright of the authors, 2016.