Science of the Total Environment xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance

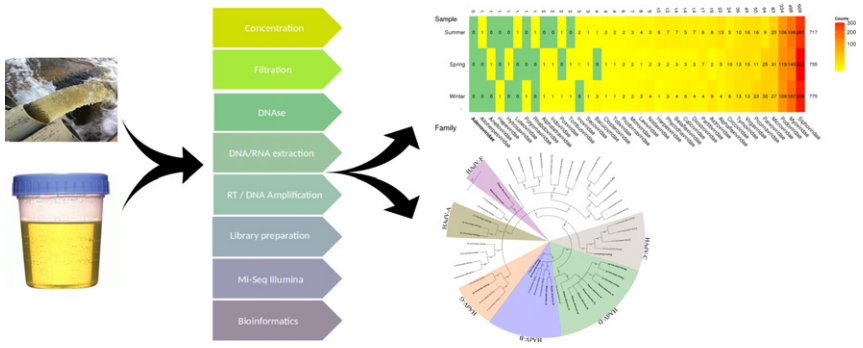
X. Fernandez-Cassi [a,*,1], N. Timoneda [a,b,1], S. Martínez-Puchol [a], M. Rusiñol [a], J. Rodriguez-Manzano [c], N. Figuerola [a], S. Bofill-Mas [a], J.F. Abril [a,b], R. Girones [a]

[a] Department of Genetics, Microbiology and Statistics, Faculty of Biology, University of Barcelona, Barcelona, Catalonia, Spain
[b] Computational Genomics Lab, University of Barcelona and Institute of Biomedicine (IBUB), University of Barcelona, Barcelona, Catalonia, Spain
[c] Centre for Bio-Inspired Technology, Department of Electrical and Electronic Engineering, Imperial College London, London, United Kingdom

HIGHLIGHTS

- Sensitive methods have been developed to study the metavirome of sewage
- The metavirome can pinpoint important pathogens circulating in the human population
- Mainly DNA viruses are excreted in urine, being JCPyV the most abundant
- The sewage metavirome is a useful tool for metagenomics in public health surveillance

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 2 June 2017
Received in revised form 23 August 2017
Accepted 24 August 2017
Available online xxxx

Editor: D. Barcelo

Keywords:
Viral metagenomics
Human adenovirus
Viral pathogens
Sewage
Next-generation sequencing

ABSTRACT

The application of next-generation sequencing (NGS) techniques for the identification of viruses present in urban sewage has not been fully explored. This is partially due to a lack of reliable and sensitive protocols for studying viral diversity and to the highly complex analysis required for NGS data processing. One important step towards this goal is finding methods that can efficiently concentrate viruses from sewage samples. Here the application of a virus concentration method based on skimmed milk organic flocculation (SMF) using 10 L of sewage collected in different seasons enabled the detection of many viruses. However, some viruses, such as human adenoviruses, could not always be detected using metagenomics, even when quantitative PCR (qPCR) assessments were positive. A targeted metagenomic assay for adenoviruses was conducted and 59.41% of the obtained reads were assigned to murine adenoviruses. However, up to 20 different human adenoviruses (HAdV) were detected by this targeted assay being the most abundant HAdV-41 (29.24%) and HAdV-51 (1.63%). To improve metagenomics' sensitivity, two different protocols for virus concentration were comparatively analysed: an ultracentrifugation protocol and a lower-volume SMF protocol. The sewage virome contained 41 viral families, including pathogenic viral species from families *Caliciviridae*, *Adenoviridae*, *Astroviridae*, *Picornaviridae*, *Polyomaviridae*, *Papillomaviridae* and *Hepeviridae*. The contribution of urine to sewage metavirome seems to be restricted to a few specific DNA viral families, including the polyomavirus and papillomavirus species. In experimental infections with sewage in a rhesus macaque model, infective human hepatitis E and JC polyomavirus were identified. Urban raw sewage consists of the excreta of thousands of inhabitants; therefore, it is a representative sample

* Corresponding author.
E-mail address: x.fernandez-cassi@ub.edu (X. Fernandez-Cassi).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.scitotenv.2017.08.249

for epidemiological surveillance purposes. The knowledge of the metavirome is of significance to public health, highlighting the presence of viral strains that are circulating within a population while acting as a complex matrix for viral discovery.

1. Introduction

In recent years, water scarcity and the application of more sustainable water reuse practices has favoured the utilisation of treated sewage for several purposes, including crop and green area irrigation, river catchment replenishment and toilet flushing. Conventional treatments applied in sewage treatment plants (STPs) are known to be less efficient for virus removal compared to faecal indicator bacteria (FIB) (Gerba et al., 1979; Pusch et al., 2005). This higher viral survival in STP treated water can represent a threat to consumers because STP effluents which contain viruses can contaminate water and food. Raw urban sewage is a complex matrix consisting of urine, faeces and skin desquamation from people. Therefore, raw sewage contains a large variety of pathogenic and commensal viruses, bacteria and protozoa excreted from thousands of inhabitants. Additionally, a high number of plant viruses pass though the human intestines. Sewage also contains other non-human inputs, which increase the diversity of this complex ecosystem. Viruses do not have conserved molecular markers, such as 16S rRNA, that are shared across all species hampering the study of viral metagenomes. However, the application of random-primer-based sequencing approaches in combination with next-generation sequencing (NGS) techniques has opened a new path for viral discovery, increasing number of new viral species described each year. Viral metagenomics applications to sewage provide excellent tools for monitoring and identifying potentially known and unknown viral pathogens that circulate among the human population, contributing to public health surveillance.

Although some viral metagenomics protocols are available for clinical samples (Kohl et al., 2015), only a few manuscripts describe the application of metagenomics approaches to analyse the viruses present in sewage (Cantalupo et al., 2011; Ng et al., 2012). Previous studies have shown that viruses prevalent in sewage are not always detected by metagenomics, suggesting that protocols should be improved to increase sensitivity. For example, in a study by Cantalupo and collaborators, showed that human adenoviruses (HAdVs) in sewage samples were barely detectable by NGS while by quantitative PCR (qPCR) the HAdV were highly prevalent.

In this study, we investigated the diversity of viruses present in raw sewage using metagenomics to test samples from three different seasons. The application of this methodology allowed for a description of the human virome and an evaluation of the sensitivity of the technique using HAdVs as a reference. HAdVs were selected because of their dual role as pathogens and specific human viral faecal indicators (Bofill-Mas et al., 2013).

With this purpose, we compared the performance of an untargeted metagenomics analysis to an adenovirus-targeted NGS assay and HAdV qPCR values. To increase the number of different viral species identified in sewage, different protocols for concentrating viruses in urban sewage were evaluated, and an efficient protocol for the analysis of viruses in sewage and other environmental samples by metagenomics has been proposed.

The application of metagenomics in different human body parts has facilitated the study of viral communities in the oral cavity (Ly et al., 2014), gut (Minot et al., 2011), respiratory tract (Willner et al., 2009), skin (Foulongne et al., 2012), blood (Sauvage et al., 2016) and cerebrospinal fluid (Perlejewski et al., 2015). Viral faecal viromes have been studied in healthy (Minot et al., 2011) and unhealthy patients (Linsuwanon et al., 2015) and in domestic animals (Mihalov-Kovács et al., 2014). Hence, the viral contribution of faeces to raw sewage seems clear. Of note, the viral communities excreted through urine remain poorly studied (Santiago-Rodriguez et al., 2015), which may be because urine has typically been considered an sterile environment. To assess the contribution of urine to the virome of raw sewage and to study the viral composition of the urine, viruses in pooled urine samples were also analysed by metagenomics.

The infectivity of the known and unknown viral species present in raw sewage was explored by intravenously inoculating a sewage sample into rhesus macaques as a potential enrichment step prior to the use of the metagenomics to examine rhesus serum samples.

Finally, a tailored protocol to analyse sewage and other environmental samples using metagenomics was proposed. Bioinformatics-specific parameters were adjusted to different levels, and new tools were tested to filter out the best set of raw reads, such as those containing the most informative sequences. These reads were combined into assembled contigs that were later used to detect the known species genomes present in the samples and relative abundances of the taxonomic groups found in the species mixture. Similarity searches also provided a basic characterisation of the pathogenic species present in the samples.

2. Materials and methods

2.1. Concentration of viral particles from tested samples

2.1.1. Concentration of viral particles from raw sewage using skimmed milk organic flocculation

Sampling points of this study are presented in Fig. 1. Three 10 L samples of raw sewage from a STP in Sant Adrià del Besós were collected in winter, spring and summer of 2013. Samples were processed 2 h after collection and stored at −80 °C until further analysis. Viral particles were concentrated using the SMF method described by Cantalupo et al. (2011). Free DNA from viral concentrates was removed, nucleic acids (NAs) were extracted, and libraries were prepared as explained in Section 2.2.

In the second protocol, a reduction in the sewage sample volume was also evaluated in order to reduce the levels of inhibitory compounds and interfering materials in the viral concentrate.

The SMF-adapted protocol used 500 mL of raw sewage that was preconditioned to a pH 3.5 and was based on the protocol described by Calgua et al. (2008). Briefly, 5 mL of a pre-flocculated skim milk solution at pH 3.5 and a conductivity superior to 1.5 mS/cm^2 was added to each sample. After 8 h of stirring, flocks were centrifuged at 8,000 ×g for 40 min, and the pellet was suspended in 4 mL of phosphate buffer [vol/vol] (0.2 M Na_2HPO_4 and 0.2 M NaH_2PO_4). The viral concentrate was kept at −80 °C until further use.

A third protocol based on ultracentrifugation was evaluated in comparison to the 500 mL SMF protocol. Two 600 mL samples of raw sewage were collected from Granollers STP. Samples were divided into two aliquots: 500 mL for processing according to the SMF-adapted protocol from Calgua et al. (2008) and 42 mL for the ultracentrifugation protocol adapted from Pina et al. (1998b). The obtained SMF and ultracentrifugation viral concentrates were filtered through 0.45 μm Sterivex filters (Millipore, Massachusetts). Free DNA was removed, NAs were extracted, and libraries were prepared as explained in Section 2.2. For both methodologies, the equivalent of 7 mL of a raw sewage sample was analysed in the final constructed libraries. The presence of HAdV was analysed by qPCR on the NA extracts as described by Bofill-Mas et al. (2006).

Please cite this article as: Fernandez-Cassi, X., et al., Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance, Sci Total Environ (2017), https://doi.org/10.1016/j.scitotenv.2017.08.249