Journal of Virological Methods 195 (2014) 194–204

Contents lists available at ScienceDirect

Journal of Virological Methods

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

Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery☆

Richard J. Hall*, Jing Wang, Angela K. Todd, Ange B. Bissielo, Seiha Yen, Hugo Strydom, Nicole E. Moore, Xiaoyun Ren, Q. Sue Huang, Philip E. Carter, Matthew Peacey

Institute of Environmental Science and Research, at the National Centre for Biosecurity & Infectious Disease, 66 Ward Street, Wallaceville, Upper Hutt 5018, New Zealand

Article history:
Received 23 May 2013
Received in revised form 26 August 2013
Accepted 29 August 2013
Available online 13 September 2013

Keywords:
Metagenomic
Virus
Purification
Enrichment

A B S T R A C T

The discovery of new or divergent viruses using metagenomics and high-throughput sequencing has become more commonplace. The preparation of a sample is known to have an effect on the representation of virus sequences within the metagenomic dataset yet comparatively little attention has been given to this. Physical enrichment techniques are often applied to samples to increase the number of viral sequences and therefore enhance the probability of detection. With the exception of virus ecology studies, there is a paucity of information available to researchers on the type of sample preparation required for a viral metagenomic study that seeks to identify an aetiological virus in an animal or human diagnostic sample. A review of published virus discovery studies revealed the most commonly used enrichment methods, that were usually quick and simple to implement, namely low-speed centrifugation, filtration, nuclease-treatment (or combinations of these) which have been routinely used but often without justification. These were applied to a simple and well-characterised artificial sample composed of bacterial and human cells, as well as DNA (adenovirus) and RNA viruses (influenza A and human enterovirus), being either non-enveloped capsid or enveloped viruses. The effect of the enrichment method was assessed by both quantitative real-time PCR and metagenomic analysis that incorporated an amplification step. Reductions in the absolute quantities of bacteria and human cells were observed for each method as determined by qPCR, but the relative abundance of viral sequences in the metagenomic dataset remained largely unchanged. A 3-step method of centrifugation, filtration and nuclease-treatment showed the greatest increase in the proportion of viral sequences. This study provides a starting point for the selection of a purification method in future virus discovery studies, and highlights the need for more data to validate the effect of enrichment methods on different sample types, amplification, bioinformatics approaches and sequencing platforms. This study also highlights the potential risks that may attend selection of a virus enrichment method without any consideration for the sample type being investigated.

1. Introduction

Since the proliferation of high-throughput sequencing technologies, the search for viruses has entered a new era (Lipkin, 2010, 2013). These technologies are capable of producing millions of sequence reads without a priori knowledge of the sample. Sequence data generated from a sample is compared to known sequence databases in order to identify viruses. Previously, virus metagenomics was accomplished by cloning and sanger-based sequencing of randomly amplified nucleic acid (Breitbart et al., 2003; Djikeng et al., 2008). Despite the small amount of sequence data produced, it was still possible to detect viruses due to either high concentrations, or some process of prior enrichment being applied that removes host cells and exogenous nucleic acid.

The relative abundance of a virus (or viral nucleic acid) in a sample, compared to that of other organisms such as bacteria or host cells (or their genomes), is a critical factor for the discovery of viruses when using metagenomics. A higher proportion of viral sequence increases the probability that (1) viral sequences will be represented in a metagenomic dataset and (2) larger contigs can be

* Corresponding author at: ESR, 66 Ward Street, Wallaceville, Upper Hutt 5018, New Zealand. Tel.: +64 4 529 0605.
E-mail addresses: richard.hall@esr.cri.nz (R.J. Hall), jing.wang@esr.cri.nz (J. Wang), angela.todd@esr.cri.nz (A.K. Todd), ange.bissielo@esr.cri.nz (A.B. Bissielo), seiha.yen@cawthron.org.nz (S. Yen), hugo.strydom@esr.cri.nz (H. Strydom), nicole.moore@esr.cri.nz (N.E. Moore), una.ren@esr.cri.nz (X. Ren), sue.huang@esr.cri.nz (Q.S. Huang), philip.carter@esr.cri.nz (P.E. Carter), matthew.peacey@gmail.com (M. Peacey).

0166-0934/$ – see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jviromet.2013.08.035

assembled, increasing the likelihood of a match in the database. It is has been shown that without some type of physical enrichment method, viruses may not be present in high enough concentrations to be detected (Daly et al., 2011). This problem is somewhat overcome as high-throughput sequencing technologies advance in both read length and sequencing depth. Nevertheless, gains in sensitivity are still possible by the application of a physical enrichment process for viruses, and may also avoid the cost of generating and analysing additional data.

In the field of viral ecology there have already been significant advances in the validation of physical enrichment methods for viruses (Duhaime and Sullivan, 2012; John et al., 2011), such as the methods used to concentrate and purify viruses from seawater for virus discovery by metagenomics (Hurwitz et al., 2013). However, only a few methodological studies have been applied to evaluate the efficiency of viral enrichment methods in metagenomics that seek to diagnose animal or human disease. A study on human liver tissue compared enrichment techniques of freeze–thaw, centrifugation and nuclease-treatment for the detection of Hepatitis C Virus using both Roche 454 and Illumina high-throughput sequencing platforms (Daly et al., 2011). The abundance of viral sequences in each treatment group was compared to results obtained by quantitative real-time PCR detection of transcripts, and the effect of each treatment method on viral genome coverage was also determined. Such studies show that physical enrichment methods do increase the sensitivity of detection for viruses in metagenomics.

For new researchers looking to perform work on human or animal samples for the purposes of detecting or diagnosing new or unexpected viruses, it can be difficult to ascertain which virus enrichment method may be applicable to a given sample type. Methods described for viral ecology studies are unlikely to be applicable. A review was undertaken of 24 published metagenomic studies that sought to describe viruses present in human or animal samples (excluding virus ecology studies) and provide details on the enrichment methods used (Table 1). Most of these studies incorporate the use of low-speed centrifugation and/or filtration to remove host cells or other micro-organisms, with a final nuclease-treatment step, where DNAse or RNAse will destroy exogenous nucleic acid but is not thought to affect nucleic acid protected by the viral capsid or envelope. Ultracentrifugation also features as a common method for the concentration of viruses from samples.

The application of virus discovery methods using metagenomics has been considered for routine use in diagnostic and reference laboratories to aid in the diagnosis of human (Svraka et al., 2010) and animal disease (Belak et al., 2013). The application of these techniques in a clinical setting will require that any virus enrichment methods are simple to perform, fast, robust, effective, standardised and do not require significant capital expenditure. It is noted that the vast majority of the published studies in Table 1 apply the simple enrichment techniques without any a priori justification for the selection of the technique. This study sought to examine the rapid and simple enrichment techniques for viruses that appear to be in routine use in the literature for diagnosing animal and human diseases, but for which the effects on metagenomic data have not been studied. This was achieved by examining the effect of these enrichment methods on the relative abundance of viruses in a metagenomic dataset derived from a simple and well-characterised artificial sample.

2. Materials and methods

2.1. Generation of an artificial sample containing bacteria, human cells and viruses

Human enterovirus 71 was cultured in human rhabdomyosarcoma cell line in Hanks MEM (Life Technologies, Carlsbad, CA, USA) supplemented with 5% foetal bovine serum (Thermofisher Scientific, Waltham, MA, USA). Human adenovirus 5 was also cultured in the human rhabdomyosarcoma cell line. Influenza A(H1N1)pdm09 was cultured in MDCK-SIAT1 cells (canine) in R-Mix (Diagnostic Hybrids, Athens, OH, USA). All virus cultures were composed of cell culture supernatant and monolayer present after freeze–thaw. *Escherichia coli* O157 was cultured in Brain Heart Infusion broth (BHI) and incubated at 37 °C overnight. Human A549 cells were cultured in DMEM (Life Technologies, Carlsbad, CA, USA) supplemented with 5% foetal bovine serum (Thermofisher Scientific, Waltham, MA, USA).

An artificial sample was formulated to consist of known amounts of *E. coli* O157, A549 human epithelial lung carcinoma cells (ATCC CCL-185), human enterovirus 71, human adenovirus 5 and influenza A(H1N1)pdm09. Aliquots of the final dilution were subjected to three freeze–thaw cycles and were frozen and stored at −80 °C.

2.2. Virus enrichment methods

Based upon a review of enrichment methods presented in Table 1, five combinations of three simple methods of enrichment were selected and performed on 1 mL aliquots of the artificial sample as follows; low-speed centrifugation in a microfuge at 6000 × *g* for 10 min at 4 °C, sterile syringe filtration at 0.45 μm, nuclease treatment using 0.1 U μL^{-1} Turbo DNAse (Life Technologies, Carlsbad, CA, USA), 0.1 U μL^{-1} RNAse One (Promega, Fitchburg, WI, USA) and 1X DNAse buffer (Life Technologies, Carlsbad, CA, USA) and incubation at 37 °C for 90 min, or combinations of these being a 2-step method (centrifugation followed by filtration), or 3-step (centrifugation, filtration then nuclease-treatment). Independent duplicates for each treatment were performed and used in all subsequent experiments.

2.3. Nucleic acid preparation

The extraction of RNA was achieved using the iPrep PureLink Virus Kit (Life Technologies, Carlsbad, CA, USA), where 400 μL of the artificial sample was extracted and eluted into 100 μL of molecular-biology grade water.

2.4. Quantitative real-time PCR assays

All real-time quantitative PCR assays (qPCR) were performed on a Stratagene Mx3000P Real-Time PCR System (Agilent Technologies, Santa Clara, CA, USA). qPCR on extracted RNA was used to quantify A549 human cells, influenza, adenovirus, *E. coli* O157 and enterovirus present in the artificial sample.

The human RNase P (RNP) gene was used as a target for the detection of human A549 cellular RNA. The nucleoprotein gene target was used for the detection of influenza A(H1N1)pdm09 RNA. Both assays were performed using the AgPath One Step RT-PCR Kit reagents (Life Technologies, Carlsbad, CA, USA) and the primers and probes for these assays have been previously described (WHO, 2011). Each 25 μL reaction contained 5 μL of nucleic acid, 12.5 μL of RT-PCR Buffer, 1 μL of 25X RT-PCR Enzyme Mix, 0.1 μM probe and 0.4 μM primers. Following an initial 30 min reverse transcription step at 50 °C and 10 min denaturation step at 95 °C, a 2-step cycling procedure of denaturation at 95 °C for 15 s with annealing and extension at 55 °C for 30 s over 40 cycles was used.

Adenovirus DNA was detected using a previously published assay (Brittain-Long et al., 2008) and the AgPath One Step RT-PCR Kit (Life Technologies, Carlsbad, CA, USA). Each 25 μL reaction contained 5 μL of DNA, 12.5 μL of RT-PCR Buffer, 1 μL of 25X RT-PCR Enzyme Mix, 0.4 μM probe and 0.5 μM primers. After an initial 10 min reverse transcription step at 45 °C and 10 min denaturation