Biologicals 43 (2015) 31–36

Contents lists available at ScienceDirect

Biologicals

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

Viral diversity in swine intestinal mucus used for the manufacture of heparin as analyzed by high-throughput sequencing

Marine Dumarest [a, 1], Erika Muth [b, 1], Justine Cheval [b], Marlène Gratigny [b], Charles Hébert [b], Léa Gagnieur [a], Marc Eloit [a, b, c, *]

[a] Institut Pasteur, Laboratory of Pathogen Discovery, Department of Virology, 28 rue du Docteur Roux, F-75724 Paris, France
[b] PathoQuest, Bâtiment François Jacob, 25 rue du Dr Roux, 75015 Paris, France
[c] Ecole Nationale Vétérinaire d'Alfort, UMR 1161 Virologie ENVA, INRA, ANSES, 7 Avenue Général de Gaulle, F-94704 Maisons Alfort, France

ARTICLE INFO

Article history:
Received 4 June 2014
Received in revised form
18 September 2014
Accepted 23 October 2014
Available online 15 November 2014

Keywords:
Virus
Heparin
Gut
Pig
Metagenomics
High-throughput sequencing

ABSTRACT

Heparin is one of the main pharmaceutical products manufactured from raw animal material. In order to describe the viral burden associated with this raw material, we performed high-throughput sequencing (HTS) on mucus samples destined for heparin manufacturing, which were collected from European pigs. We identified Circoviridae and Parvoviridae members as the most prevalent contaminating viruses, together with viruses from the *Picornaviridae*, *Astroviridae*, *Reoviridae*, *Caliciviridae*, *Adenoviridae*, *Birnaviridae*, and *Anelloviridae* families. Putative new viral species were also identified. The load of several known or novel small non-enveloped viruses, which are particularly difficult to inactivate or eliminate during heparin processing, was quantified by qPCR. Analysis of the combined HTS and specific qPCR results will influence the refining and validation of inactivation procedures, as well as aiding in risk analysis of viral heparin contamination.

1. Introduction

Heparin sodium is the purified sodium salt of heparin, a high molecular weight polysaccharide derived from porcine intestinal mucosa. Heparin sodium, or more frequently its derivative, low-molecular-weight heparin (LMWH), is used as a class of anticoagulant medications [1]. Porcine heparin is prepared either from porcine intestinal mucosa or from whole minced gut. One pig is necessary to manufacture three doses of purified heparin or one dose of LMWH, and around 100 tons of heparin are manufactured every year [2]. The worldwide demand for both heparin sodium and LMWH has increased over the last few years, and currently, more than 20 million pigs worldwide are used for its manufacture each year. The raw material is likely to be rich in enteric viruses which are generally excreted at high titers, and which are often resistant to many physical and chemical treatments, therefore any recipient human patients could theoretically be exposed to porcine viruses. Moreover, increasing knowledge of the porcine enteric virome has so far uncovered greater viral diversity than previously thought [3–7,22]. The heparin manufacturing process involves numerous steps [1], each of which must undergo verification for their efficacy in either inactivating or removing viruses, according to current regulations. To help build contamination risk analyses, and to establish which viruses should be monitored, investigations must be undertaken to determine those viruses likely to contaminate the raw material, and their respective viral loads.

High-throughput sequencing (HTS) techniques efficiently detects viruses present in biological fluids (reviewed in Ref. [8]) and are increasingly being used in medical diagnosis [9,10] or for screening biological materials [11,12]. We have developed a pipeline, from sample preparation to bioinformatics [13], able to identify known, as well as new viruses [14,15] and have recently demonstrated its use in evaluating the viral burden of fetal calf serum and trypsin used in cell culture [16]. Here we describe the analysis of viruses present upstream of the heparin manufacturing process. We show that a diverse range of both known and unknown

Abbreviations: Gc/mL, genome copies/mL; HTS, high-throughput sequencing; PAV A, porcine adenovirus A; PAV B, porcine adenovirus B; PBoV, porcine bocavirus; PCV, porcine circovirus; PCV1, porcine circovirus type 1; PCV2, porcine circovirus type 2; PPV1-6, porcine parvovirus 1 to 6; PPV7, putative porcine parvovirus 7; PRV-A, porcine group A rotavirus.

* Corresponding author. Institut Pasteur, Laboratory of Pathogen Discovery, Department of Virology, 28 rue du Docteur Roux, F-75724 Paris, France. Tel.: +33 1 44 38 92 16.
E-mail address: marc.eloit@pasteur.fr (M. Eloit).

[1] MD and EM contributed equally to this work.

http://dx.doi.org/10.1016/j.biologicals.2014.10.004
1045-1056/

viruses are present in this raw material, and discuss the impact of these results on establishing requirements for future viral validations.

2. Methods

2.1. Samples and preparation

A total of 10 mucus pool samples were collected in Europe, reflecting the diversity of the slaughterhouses utilized by a single manufacturer over several months. Each pool represented several hundreds of pigs and corresponded to the raw material used for pure heparin sodium manufacturing. 1 mL of each mucus sample was mixed with 9 mL of N-acetyl cysteine (Merck Millipore, Billerica, MA) at a concentration of 100 mg/mL, vortexed, centrifuged for 20 min at 4000 rpm at 4 °C. The supernatants were filtered through a 0.22 μm filter and the virus particles of each pool were independently concentrated by ultracentrifugation for 2 h at 100,000 *g* through a cushion of 30% w/v sucrose. The pellet was resuspended in 150 μL of water and treated with a cocktail of nucleases adapted from metagenomic study of gut contents to digest non particle-protected nucleic acids (Turbo DNase (final concentration, 20 U/ml; Ambion) and RNase A (final concentration, 0.1 mg/ml; Fermentas) at 37 °C for 30 min) [17]. Enzymes were inactivated with a final concentration of 3 mM EDTA and heating at 10 min at 65 °C. The virus particles-associated genomes contained in 80 μL of each mucus pool sample were extracted with the Cador Qiagen Pathogen minikit (Hilden) and then amplified by the bacteriophage phi29 polymerase based multiple displacement amplification (MDA) assay using random primers. This technique allows DNA synthesis from DNA samples, and also from cDNA fragments from viral genomes previously colligated prior to Phi29 polymerase-MDA [18]. A mix with 4 μl of nucleic acids, 0.5 μl of primer (50 μM) and 0.5 μl of dNTPs (10 mM) was incubated at 75 °C for 5 min and cooled on ice for 5 min. Then, 5 μl of enzyme mix were added. This enzyme mix was composed of 2 μl of 10× RT Buffer for SSIII (Invitrogen Inc. Saint Aubin, France), 4 μl of 25 mM $MgCl_2$, 2 μl of 0.1 M DTT, 1 μl of 40 U/μl RNaseOUT (Invitrogen Inc., Saint Aubin, France), 1 μl of SuperScript III reverse transcriptase (Invitrogen Inc.). The final mix was incubated at 25 °C for 10 min, then at 45 °C for 90 min and finally at 95 °C for 5 min. The two following steps (ligation and MDA) were performed with the QuantiTect® Whole Transcriptome kit (Qiagen) according to the manufacturer's instructions. Each of the ten samples provided concatemers of high molecular weight DNA at a concentration close to 1 μg/μL that were pooled before sequencing. Sample extraction and random amplification procedures were carefully performed to prevent cross-contamination, using the best precautionary PCR standards.

2.2. HTS and bioinformatic analysis

Reads were generated on an Illumina® HiSeq-2000 sequencer (DNAVision, Gosselies, Belgium) with a sequencing depth of 2.4×10^8 paired-end reads of 101 nt in length. Sequences were trimmed and filtered according to their quality score. Sequencing library preparation may introduce residual sample cross-contamination. After porcine genome sequence subtraction (susScr3, SGSC Sscrofa10.2 – NCBI project 13421, GCA_000003025.4, WGS AEMK01) with Cushaw2 and BlastN, reads were assembled in contigs using CLC Genomics Assembly Workbench (Cambridge, USA), and contigs and singletons were assigned a given taxonomy using the Blast algorithm. Criteria for taxonomic assignation have been described previously [16]. Sequences of the main contigs are available upon request.

2.3. PCR

Quantitative PCR was used to quantify virus loads for the known or candidate non-enveloped viruses identified in this study. SYBR green qPCR amplification was carried out in 20-μl reaction volumes that contained 2 μl of DNA, 1X Master Mix, and 500 nM each of the forward and reverse primers respectively (Table 2) (LightCycler 480 SYBR Green I Master, Roche Diagnostics, Meylan, France). qPCR analyses of all samples were performed in duplicate, and were conducted as indicated in Table 2 using the following primers: generic primers for known viruses (PCV1/2 and porcine bocavirus), or specifically designed primers based on a major contig, for unknown viruses (PPV7). Calibration curves were generated using a purified amplicon at known concentrations as control standards.

2.4. Role of the funding source

The study sponsors were not involved in study design, data collection, data analysis, data interpretation, or the writing of the report. The corresponding author had full access to all data in the study and had final responsibility for the decision to submit for publication.

3. Results

3.1. Description of the viruses present in pig mucus

Fig. 1 depicts the proportion of reads corresponding to sequences that closely match known porcine viruses. The vast majority of viruses were found to be non-enveloped viruses, except for a few reads related to the *Herpesviridae* family, and reads from endogenous retroviruses that were likely to originate from contaminating porcine DNA. Members of the *Parvoviridae* family represented 76.3% of total viral reads, and within this group members of the bocavirus genus represented 79.5% of these *Parvoviridae* reads, followed by *Partetravirus* genus members (13.8%), as shown in Fig. 2. Members of the *Circoviridae* family represented 16.3% of the total viral reads, which were mostly composed of PCV2 viruses (98.6%), while the remaining reads mapped to PCV1 and Po-Circolike virus 22 (data not shown) [6]. Sequences of the NIH-CQV virus, a known contaminant of Qiagen extraction columns [19,20] were also identified and discarded. Other frequent reads (2.49%) were from *Picornaviridae* viruses, and more specifically from the newly described genus proposed as *Pasivirus* [3] (accounting for 78% of these reads, data not shown). Other viral families such as *Picobirnaviridae, Reoviridae* (mainly rotavirus A to C), *Adenoviridae* (mainly PAV A and B), *Astroviridae*, and *Caliciviridae* (mainly porcine sapovirus), were also represented, but at much lower frequencies.

Putative new viral species were also identified. To address the study aim, we focused on those viruses that could be challenging to remove during the manufacturing process, because they belong to families known for either their physical resistance, their small size, or both (Table 1). Detailed results are presented in Supplementary Table S1. We identified potential novel viral species in the *Astroviridae*, *Caliciviridae*, *Circoviridae*, *Parvoviridae*, and *Reoviridae* families. The most frequent reads corresponded to members of the *Parvoviridae* family, and were distantly related (around 64% amino acid identity) to known parvoviruses. The most closely related was the Eidolon helvum parvovirus 2, an unclassified member of the partetravirus genus found in frugivore bats of Africa [21]. This suggests the presence of at least one new porcine parvovirus species that we have tentatively named porcine parvovirus 7. Fig. 3 shows that PPV7 clusters with Eidolon helvum parvovirus 2,