Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
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 polysaccharide derived from porcine intestinal mucosa. 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 polysaccharide derived from porcine intestinal mucosa. 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 polysaccharide derived from porcine intestinal mucosa.
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 (Turbto 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 MgCl2, 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 concatamers 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, GosseIs, 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 preparation may introduce residual sample cross-contamination. After porcine genome sequence subtraction (susScr3, Sscrofa10.2 – NCBI project 13421, GCA_000003025.4, WGS AEMK01) with Cusshwaw2 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 Porcicoliike 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 Pichobirnaviridae, 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 paroviruses. The most closely related was the Eildon 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 parovirus species that we have tentatively named porcine parovirus 7. Fig. 3 shows that PPV7 clusters with Eildon helvum parvovirus 2,
4. Discussion

We describe here the viral burden of pig intestinal mucus, the most frequently used raw animal material for the manufacture of a biological product, heparin. To our knowledge, this is the first broad viral analysis of such material. We show here that numerous viral sequences are present in the raw pig intestinal mucus, as expected for samples directly derived from gut contents. As the material utilized for high-throughput sequencing were nuclease-treated pellets resulting from ultracentrifugation, it is likely that sequences obtained correspond to whole virus particles, even if we cannot totally exclude that non-encapsidated nucleic acids protected within aggregates might have also influenced results.

Total viral content was dominated by non-enveloped viruses, typical of enteric viruses. We identified members of the Circoviridae and Paroviridae families as the major mucus-contaminating viruses. The fecal pig virome, which in theory should be similar to the mucosal virome, was examined recently via HTS and our findings were similar in relation to the main virus families identified. Nevertheless, the proportion of each virus family generally differed, which could perhaps be due to variation between animals, where studies were conducted on individual pigs from one or a limited number of herds [5,6]. It should be emphasized that most of the identified viruses were not porcine pathogens. For example, whilst PCV2 (the main detected species) is responsible for the post-weaning multisystemic wasting syndrome [23], PCV1 seems to be non-pathogenic. Porcine bocaviruses are diverse and have not yet been associated with disease [24,25]. These are interesting findings, as most testing guidelines for the viral safety of biological products are dominated by the search for porcine pathogens. This is evidently due to bias as veterinary virology is dominated by research on animal diseases. Indeed, most of the zoonotic viruses are dominated by the search for porcine pathogens. This is evidently due to bias as veterinary virology is dominated by research on animal diseases. Indeed, most of the zoonotic viruses are either weakly, or not at all pathogenic in their animal reservoir. Therefore we should remain cautious about predicting the impact of such “non-pathogenic” animal viruses on human health. Moreover, we mainly detected positive-ssRNA or –ssDNA viruses, which are known to harbor marked capabilities in adapting to new hosts following successful initial cross-
transmission events. On the other hand, we did not detect viruses known to be transmissible to humans, such as influenza [26], HEV [27] and EMCV [28], while the zoonotic status of some of the identified viruses (rotavirus [29] and norovirus [30]) are still the subject of fierce debate.

The level of sensitivity of our pipeline is close to that of PCR for known viruses as shown previously for a depth of sequencing close to 8 million reads per sample [13] and confirmed recently for a higher depth of sequencing similar to that used in this study, which in addition allows for a better genome coverage [31]. So, it seems unlikely that a high load of a virus able to challenge the drastic manufacturing process of heparin could escape detection. The pipeline has also been shown to detect viruses very distant from known species (this paper and [3]), but it remains indeed possible that a virus very far from those already present in databases might escape such detection.

The number of NGS reads is not proportional to the relative abundance of viral genomes, as the different genome types (single/ double stranded DNA/RNA) are amplified differently. Also, the coverage of the viral genomes is generally not uniform [13]. So, it seems unlikely that a high load of a virus able to challenge the drastic manufacturing process of heparin could escape detection. The pipeline has also been shown to detect viruses very distant from known species (this paper and [3]), but it remains indeed possible that a virus very far from those already present in databases might escape such detection.

The number of NGS reads is not proportional to the relative abundance of viral genomes, as the different genome types (single/ double stranded DNA/RNA) are amplified differently. Also, the coverage of the viral genomes is generally not uniform [13]. So, it is currently impossible to estimate virus loads from NGS results. Due to the study’s objective and the resultant viral diversity, we decided to focus quantitative analysis on a subset of those viruses which may be especially resilient to removal or inactivation during the manufacturing process i.e. Circoviridae and Paroviridae. Both are very resistant to physical and chemical inactivation, and in addition, are the smallest of vertebrate viruses (17–24 nm and 18–26 nm respectively), and thus are the most difficult to clear by nanofiltration. Among the Paroviridae, we chose two species: porcine bocavirus, representing 79.5% of Paroviridae reads, and the new PPV7 virus, to model unknown viruses which would not have been detected using current PCR methods. Circoviruses and the two paroviruses were present in 9/10 batches (PPV7 was not detected in batch 10). Viral loads were high and remarkably similar between batches, which is probably a consequence of frequent shedding and the large size of the tested pools (several hundred pigs), which probably averages out viral loads. The ten mucus batches contained between 7.6 and 8.5 log gc/mL of several non-enveloped small DNA viruses, which represents severe challenges for downstream purification processes.

Animals are sourced worldwide for heparin purification, including animals from North America and China. Consequently, as these mucus samples were collected from European herds, results may not be representative of all mucus sources. Nevertheless, it is likely that certain resident viruses represent a viral profile characteristic of this animal species. This analysis did not take into account geographical sources of variation, which are further

Table 2

| Virus Family | Primer Sequence (5’-3’) | Amplicon size (nt) | Cycling conditions | Reference |
|--------------|--------------------------|-------------------|--------------------|-----------|
| Circovirus 1 et 2 | Forward TGGCCCGCCATTTTTGATT ReReverse CAGGCGGCAGGCGGTTGAG | 72 | 45 cycles 95 ºC: 10 s 58 ºC: 5 s 72 ºC: 10 s | [34] |
| Bocavirus | Forward GTACCAGATCTGATGATGACAC ReReverse AAAGACCCAARAAATATT | 231 | 45 cycles 95 ºC: 10 s 47 ºC: 5 s 72 ºC: 10 s | [25] |
| PPV7 | Forward TGGTCGTGATGATGATGGG ReReverse CGCAGAGAAAGCCAAACAAG | 104 | 45 cycles 95 ºC: 10 s 56 ºC: 5 s 72 ºC: 10 s | |

Fig. 1. Viral reads derived from pig mucus corresponding to known viruses Ratio of viral reads for each virus family to the total number of unique (non-duplicated) viral reads closest to a known virus species derived from the sample (456,437 reads).

Fig. 2. Viral reads derived from pig mucus matching to known members of the Paroviridae. Ratio of viral reads for each genus to the total number of unique (non-duplicated) viral reads closest to a known virus species from the Paroviridae family (348,363 reads).
complicated by the multiplicity of pig strains. Neither did it examine the impact that any enteric viral diseases could have on viral excretion. Coronaviruses, such as the porcine epidemic diarrhea virus (PEDV), transmissible gastroenteritis virus (TGEV), and rotaviruses (PRV-A), are major porcine viruses causing enteric disease [32,33]. Even so, the Reoviridae members (including rotaviruses) were poorly represented in the viral count (0.9%), and no coronaviruses were identified. Inclusion of a herd with acute viral diarrhea would have most likely have modified the mucosal viral composition.

Our results should help to define guidelines for the appropriate validation of procedures for the inactivation of pertinent resistant viruses, like parvoviruses and circoviruses, the two main adventitious viruses revealed here. To validate these inactivation processes, Porcine parvovirus (PPV) or any other Parvoviridae member would represent a relevant reference virus. Use of circovirus would also aid in the validation of the more challenging nanofiltration steps. The choice of enveloped viruses classically used in the validation of manufacturing processes appears to be of lower interest compared to the risks of raw material contamination assessed here.

Evaluation of the probability of survival of viruses in the final product would necessitate to subtract the reduction factor of validated steps of the process from the load of viruses upstream of the process. This is outline the scope of the paper as this would necessitate to know not only the viral titers in the mucus (this study), but also the amount of mucus used for the manufacture of each dose, and the validated reduction factors of the manufacturing process.

Currently, mucus samples do not undergo viral testing prior to processing. In any case, the assays would have been uninformative, as all mucus samples contain viruses, and moreover, there are neither bio- nor molecular assays available for several virus types. Heparin safety thus relies on efficient inactivation and/or removal capabilities during processing. Using a combination of NGS analysis and quantitative PCR techniques, it is now feasible to characterize the viral burden of such raw materials. The resulting in-depth data of viral species and loads would then guide the selection of viruses used to validate inactivation processes, and could also be used to build risk analyses needed for the release of biological products on the market.

Acknowledgments

This work received financial support from the BioSecure Foundation, Paris, Virus-Biol-HTS and the Labex IBEID. We thank Rachel Pert for the English correction of the manuscript.

### Appendix A. Supplementary data

Supplementary data related to this article can be found at [http://dx.doi.org/10.1016/j.biologicals.2014.10.004](http://dx.doi.org/10.1016/j.biologicals.2014.10.004).

### References

[1] Linhardt RJ, Gunay NS. Production and chemical processing of low molecular weight heparins. Semin Thromb Hemost 1999;25(Suppl. 3):5–16.

[2] Liu H, Zhang Z, Linhardt RJ. Lessons learned from the contamination of heparin. Nat Prod Rep 2009;26:313–21. [http://dx.doi.org/10.1039/b819859a](http://dx.doi.org/10.1039/b819859a).

[3] Sauvage V, Ar Gouilh M, Cheval J, Muth E, Pariente K, Burguieres A, et al. A member of a new picornaviridae genus is shed in pig feces. J Virol 2012;86:10036–46. [http://dx.doi.org/10.1128/JVI.00046-12](http://dx.doi.org/10.1128/JVI.00046-12).

### Table 3

| Mucus lot | PCV1-2 | Bocavirus | PPV7 |
|-----------|--------|-----------|------|
| 1         | 8.3a   | 8.4       | 8.4  |
| 2         | 8.2    | 8.2       | 8.2  |
| 3         | 8.1    | 8.1       | 8.1  |
| 4         | 8.1    | 8.2       | 8.3  |
| 5         | 8.7    | 8.4       | 8.2  |
| 6         | 8.4    | 8.2       | 8.2  |
| 7         | 7.6    | 7.9       | 7.9  |
| 8         | 7.7    | 7.8       | 8.4  |
| 9         | 7.8    | 7.9       | 6.9  |
| 10        | 8.5    | 8.0       | Neg  |

*a Log genome copies per mL mucus, average of two replicates.
