Virologisches Institut
der Vetsuisse-Fakultät Universität Zürich

Direktor: Prof. Dr. Cornel Fraefel

Arbeit unter wissenschaftlicher Betreuung von
PD Dr. Claudia Bachofen

A next generation sequencing approach for the diagnosis of outbreaks of emerging, mutated or novel viral pathogens

Inaugural-Dissertation

zur Erlangung der Doktorwürde der
Vetsuisse-Fakultät Universität Zürich

vorgelegt von
Jakub Kubacki
Tierarzt
von Piła, Polen

genehmigt auf Antrag von
PD Dr. Claudia Bachofen, Referentin
Prof. Dr. Volker Thiel, Korreferent

2018

CONTENTS

ZUSAMMENFASUNG **3**
ABSTRACT **4**
VIRUS GLOSSARY **5**
1. INTRODUCTION **6**
2. AIMS **9**
3. MATERIALS AND METHODS **10**
3.1 General schedule 10
3.2 Animals, samples, sources of the samples 11
3.3 Spiked-in viruses for tests 12
3.3.1 Bovine viral diarrhea virus (BVDV) 12
3.3.2 Bovine Herpesvirus 1 (BoHV-1) 12
3.3.3 Swine influenza A virus (SIV) 12
3.3.4 Porcine Rotavirus A (PRV-A) 13
3.3.5 Porcine Parvovirus (PPV) 13
3.3.6 Transmissible Gastroenteritis virus (TGEV) 13
3.4 Homogenization 13
3.5 Filtration 14
3.6 Nuclease treatment 14
3.7 Nucleic acid isolation 15
3.8 Reverse transcription and second strand synthesis 15
3.9 Random amplification 16
3.10 Library preparation and NGS setup 17
3.11 Testing of the virome protocol using spiked samples 18
3.12 Testing the virome protocol with clinical samples 19
3.13 Testing the sensitivity of the protocol 19
3.14 Full-length sequencing of Hepatitis E virus (HEV) 19
3.15 Whole genome sequencing and Swiss SIV surveillance program 19
3.16 Data analysis 20
4. RESULTS **21**
4.1 Homogenization 21
4.2 Filtration 22
4.3 Nuclease treatment 24
4.4 Nucleic acid isolation 26
4.5 Reverse transcription and second strand synthesis 27
4.6 Random amplification (SISPA) 28

4.7 Library preparation kits 30
4.8 Testing the virome protocol 32
4.9 Clinical samples 34
4.10 Sensitivity of the protocol 35
4.11 Full-length sequencing of Hepatitis E virus (HEV) 36
4.12 Whole genome sequencing and Swiss SIV surveillance program 36
5. DISCUSSION 38
REFERENCES 46

ZUSAMMENFASUNG

Im Zeitalter von Globalisierung, Klimawandel und weltweitem Handel wird die Tiergesundheit in zunehmendem Masse durch Krankheiten gefährdet, welche durch sogenannte «emerging» Pathogene verursacht werden, sprich Krankheitserreger, die neu auftreten oder in neuen Varianten wieder-auftreten. Klassische Nachweismethoden wie ELISA oder PCR müssen gezielt eingesetzt werden und sind spezifisch für einen einzelnen bekannten Erreger oder eine umschriebene Gruppe verwandter Pathogene. Darum wird heute in Fällen, wo Standard-Methoden nicht zum Ziel führen, in zunehmendem Mass die next generation sequencing (NGS) Methode angewendet. Da dabei die gesamten in einer Probe vorliegenden Nukleinsäuren sequenziert werden, können auch alle darin enthaltenen Viren, das sogenannte Virom, bestimmt werden. Allerdings gibt es kaum NGS-Protokolle, die auf die Bedürfnisse der Veterinär-Diagnostik zugeschnitten sind. Darum präsentieren wir in dieser Arbeit eine Methode für die Aufbereitung tierischen Probenmaterials für NGS Untersuchungen in der Veterinär-Virologie.

Um ein NGS Protokoll speziell für die Veterinär-Diagnostik zu etablieren sollten einige Punkte beachtet werden. Die Methode sollte sensitiv, schnell und kosten-effizient sein und dabei für eine Vielzahl von Probenmaterialien und Spezies verwendbar sein. In diesem ersten Schritt haben wir uns auf die Proben-Aufbereitung konzentriert. Verschiedene Möglichkeiten bezüglich wichtiger Parameter wie Porengrösse der Filter, Nukleasen-Behandlung, Extraktionsmethode und Amplifikation wurden mittels gespikter Proben (künstlich mit bekannten RNA und DNA Viren versetztem Material) getestet und verglichen. Die Effizienz jedes Schrittes für die relative Anreicherung von Viruspartikeln wurde mittels spezifischer real-time (RT-)PCRs und DNA Messungen im Fluorometer bestimmt.

Die Analyse erster Sequenzdaten zeigte, dass die relative Anreicherung von Viruspartikeln durch eine Kombination von Filtration, Zentrifugation und Nukleasen-Behandlung die Nachweisbarkeit verschiedenster Viren massiv verbesserte. In einem nächsten Schritt wurde das Verfahren anhand archivierter klinischer Proben verschiedenster Spezies getestet, die vorgängig mit für bestimmte Viren getestet worden waren. In allen Fällen bestätigten die NGS Daten die Resultate der spezifischen Untersuchungen. Bei einigen Proben konnte sogar die vollständige virale Genom Sequenz bestimmt werden; so im Fall von Hepatitis E Virus, porcinem Circovirus-2 und allen 8 Segmenten eines Influenzavirus-Genoms. In einigen Fällen, wo die spezifischen Untersuchungen zu keinem Resultat führten, wurden mittels NGS neue oder unerwartete Viren entdeckt, wie zum Beispiel Torque-Teno sus Virus im Hirn eines Schweines mit neurologischen Symptomen. Abschliessend kann gesagt werden, dass die vorliegende Arbeit eine wichtige Basis für die Einführung eines diagnostisch anwendbaren NGS-Protokolls für die Veterinär-Virologie bildet.

Schlagworte: Probenaufbereitung, metagenomische Sequenzierung, next generation sequencing, Virusnachweis, Virom, Diagnostik, Veterinärmedizin

ABSTRACT

Animal health in the wake of intensive globalization, climate change and globetrotting is challenged by an increasing number of viral diseases often caused by emerging, re-emerging or novel viral variants. Classical methods such as enzyme-linked immunosorbent assay (ELISA) or polymerase chain reaction (PCR) are limited to circumscribed groups of known pathogens, require careful selection and a targeted approach. Therefore, in cases of disease outbreaks where standard routine diagnostics fail to identify the infective agent, next generation sequencing (NGS) technology enables to identify and discover viruses without the need to target a specific infectious agent because all viruses present in a sample are sequenced, the so-called virome. However, there is still a lack of applicable NGS protocols fitted to veterinary diagnostics purposes. Here, we present an animal sample preparation protocol for high-throughput metagenomic virus sequencing.

In order to establish a NGS protocol applicable in veterinary diagnostics, several points have to be attained. The protocol should be sensitive, fast, cost-efficient and adaptable to diverse samples materials and species. In a first step, we focused on the sample preparation. Parameters such as filter pore size, nuclease treatment, extraction method and amplification cycle numbers were tested and compared using samples spiked with RNA and DNA viruses. Effectiveness of each method for removal of host nucleic acid and preservation of viral genomes was measured by specific real-time (RT)-PCR and DNA concentration measurements on Qubit Fluorometer.

Analysis of the sequence data showed that enrichment for virus particles such as filtration, centrifugation and nuclease treatment as well as amplification have significant influence on the number of viral genomic sequences detected. In a second step, the enrichment protocol was tested in further sequencing runs, using clinical samples collected at different clinics of the Tierspital (University of Zurich) from several species with or without clinical symptoms. In all cases, NGS data reflected nicely the results of previously performed specific (RT)-PCR results. In addition, it provided full-length genome sequences of two hepatitis E virus genotype 3 (HEV-3) isolates, porcine circovirus type 2 (PCV-2) and all 8 segments of swine influenza virus (H1N1). In other cases, where previous tests were negative, new or unexpected viruses where detected such as Torque teno sus virus in the brain of a pig with neurological signs. In conclusion, the present study sets a basis for the development of a diagnostically applicable metagenomic sequencing protocol in veterinary medicine.

Keywords: Sample preparation, Metagenomic sequencing, next generation sequencing, Virus detection, Virome, Diagnosis, Veterinary Medicine

VIRUS GLOSSARY

BoHV-1	Bovine herpesvirus 1
BVDV	Bovine viral diarrhea virus
HCV	Hepatitis C virus
HEV	Hepatitis E virus
HIV	Human immunodeficiency virus
OvHV-2	Ovine herpesvirus 2
PCV-2	Porcine circovirus 2
PPV	Porcine parvovirus
PRV-A	Porcine rotavirus A
SIV	Swine influenza virus
TGEV	Transmissible gastroenteritis virus
TTV	Torque teno virus

1. INTRODUCTION

More than 50% of emerging diseases or infections are of zoonotic origin and involve a broad range of infectious agents: viruses, bacteria, fungi and parasites. A microbe is described as emerging if it is newly recognized, newly evolved or has occurred previously but shows increase in incidence, expansion into a new ecological niche or geographical zone, host or vector range (World Organization for Animal Health). Pathogens causing zoonosis such as Swine Influenza virus, porcine Hepatits E virus, Nipah virus, Staphylococcus aureus and Streptococcus suis have for example been reported in pigs (Smith TC et all., 2011). Furthermore, new viruses or viral variants have been detected in pigs such as the potentially zoonotic Ndumu virus in Uganda (Masembe et al., 2012) or a novel porcine H1N2 Influenza virus in Germany (Lange et al., 2013). However, not only pigs but many other domesticated and wild animal species harbor potentially zoonotic viruses and intensive globalization, climate change and globetrotting facilitate the spread of viruses and their vectors across the world. This situation sets many challenges to global and national veterinary health organizations. An important key to managing outbreaks and taking appropriate measures for prevention of further spread is having efficient novel diagnostic approaches at hand for the detection and identification of "new'' or unexpected viruses (Belak et al., 2013).

The rapid development of novel molecular methods enables to identify transmissible pathogens through the detection and identification of nucleic acid. Molecular methods such as nucleic acid hybridization, polymerase chain reaction (PCR) or nucleic acid sequencing, allowed to improve investigations of nucleotide sequences and are crucial for biological research. The major breakthrough in DNA sequencing was in 1977, when Frederick Sanger and colleagues developed a sequencing method, based on chain-termination technique (Sanger et al., 1977). Sanger sequencing became and remained the most widely-used technology to sequence DNA. The first-generation sequencing DNA machines based on the Sanger method generated reads of maximum length of 1000bp. Therefore, to analyze longer DNA fragments, sequencing of overlapping DNA fragments was necessary. In 2001, i.e., after 11 years of using the Sanger method-based sequencing as a main tool and assembling multiple cloned sequences into a contiguous long sequence in silico, the human genome was published (Lander et al., 2001). However, the limitations of the Sanger method, such as throughput, speed and low quality in the first bases, promoted researches to develop radically different sequencing methods, that

would bring fast and accurate genome information. These attempts resulted in the development of next generation sequencing (NGS) technology. NGS is a relatively novel technique that provides high speed and high throughput of enormous numbers of sequences. Additionally, it offers the possibility to sequence up to 96 different samples in a single run. Although the read length is shorter, i.e., 75-300bp comparing to 700-1000bp standard fragment length in Sanger method, compilation of the short fragments allows for sequencing of large genomes without requiring specific primers. The cost of whole genome sequencing has therefore been lowered dramatically. At the beginning of NGS, there were four different platforms available on the market: (1) 454 (Roche), that used the pyrosequencing method, (2) Ion Torrent (Life Technologies), that used semiconductor sequencing and was independent of using any optical device, (3) SOLiD, which based on sequencing by ligation and two base coding; and (4) - Illumina/Solexa, that based on sequencing with reversible terminators (van Dijk et al., 2014).

It has to be emphasized that the choice of the NGS platform should be based on needs and expectations of the user and should consider aspects such as costs, sequencing speed, read length, throughput and accuracy.

In the wake of decreasing sequencing costs, NGS is becoming progressively exploitable in the veterinary diagnostic field. Classical methods such as enzyme-linked immunosorbent assay (ELISA) or PCR are limited to circumscribed groups of known pathogens, require careful selection and a targeted approach. Using NGS technology, identification and discovery of viruses is possible without the need to target a specific infectious agent – all viruses present in a sample are sequenced. NGS is therefore ideal to study the so-called virome, the entirety of viruses present in a specific environment. However, NGS will not replace conventional diagnostic methods but it may be helpful in cases of disease outbreaks when standard routine diagnostics fail to identify the infective agent. Thus, for instance, metagenomic NGS approaches allowed to detect and identify Schmallenberg virus (Hoffmann et al., 2012). Additionally, the detection of a novel potentially zoonotic coronavirus in bats from China (Wu et al., 2012) and the involvement of Torque teno and boca-like parvovirus in PRRS (Blomstroem et al., 2010) was possible by applying NGS technology. Furthermore, this method has appeared to be extremely helpful in rapid identification of new viral variants, e.g. Bluetongue (Boyle et al., 2012) or reassorted viruses, e.g. avian influenza (Van Borm et al., 2012). Moreover, NGS is widely used in HIV quasispecies analysis (Johanson and Geretti, 2010). Finally, use of NGS was reported (Feng et al., 2008) for detection of new tumor viruses i.e., Merkel Cell polyomavirus. Since NGS allows to sequence full-length viral genomes, the

genome diversity and evolution can be investigated (Tebit et al., 2011). Moreover, NGS appears to be a useful tool in investigating virus-host and virus-other infectious agents interactions, allowing better prediction of the disease outcome, improved treatment, or identification of novel therapeutic strategies.

Comparing NGS to classical diagnostics methods, there are also a couple of downsides such as: lower sensitivity than PCR in viral genome detection, time consuming sample preparation, or need of immense computing resources and frequent bioinformatics support for data analysis. An important aspect in applying NGS to veterinary diagnostic is the choice and preparation of an adequate sample (Thomas et al., 2012). While some samples, such as feces, are abundant and easy to collect, not all viruses are excreted by this route. For example, respiratory infectious agents are most efficiently detected in bronchoalveolar (BAL) fluid that can be collected only after sedation in clinical conditions. Another important aspect that needs to be considered is sample contamination with high amount of host RNA and DNA, which also will be subjected to sequencing. Therefore, material containing a high number of host-cells such as tissue or blood require special enrichment for viral particles, or higher sequencing capacity to detect virus when viral load is low. This enrichment may encompass procedures such as filtration, centrifugation or nuclease treatment. These methods, however, have to be performed cautiously and carefully established, as too harsh treatment may cause removal of some viruses, and therefore, lower diagnostic quality of the sample.

2. AIMS

The aim of this project is to make a first step towards the development of a diagnostically applicable NGS method, that can be implemented in veterinary virology when routine diagnostic tools fail. There are already protocols available for virome analysis, e.g. from environmental samples or in human cases. However, the high diversity of different sample material (tissue, blood, all sorts of secretions and excretions) and host species diversity as well as limited financial means in veterinary diagnostics demand for a special approach. Being tailored to the veterinary diagnostic requirements, this method should enable quick, comprehensive, species-independent but also cost-efficient identification of the viral pathogens present, in case of disease outbreaks, in undiagnosed cases where other diagnostics measures failed, or for surveillance of a population. In this first step emphasis is placed on different techniques for sample preparation and particularly enrichment for virus particles as they have significant influence on the efficiency and breadth of virus detection and depend heavily on the type of sample to be used (Conceicao-Neto et al., 2015).

3. MATERIALS AND METHODS

3.1 General schedule

In order to establish a protocol that could be applied for metagenomic sequencing of animal samples, the following steps were differentiated: (I) samples homogenization, (II) filtration, (III) nuclease treatment, (IV) DNA and RNA extraction, (V) reverse transcriptase and second strand synthesis, (VI) DNA random amplification, (VII) library preparation (Fig.1). Initially, each step was tested with different kits, materials and conditions to determine the impact of different sample treatments on the recovery of viral genomes. To determine the effects on viral genomes, all the samples were measured by quantitative PCR (qPCR) on the Applied Biosystems™ QuantStudio™ 7 Flex Real-Time PCR System. To find the most efficient way of reducing the sample contamination with DNA originating from host, bacteria or other cell type, DNA and RNA quantitation have been measured using Qubit™ dsDNA BR Assay Kit, Qubit™ dsDNA HS Assay Kit and The Qubit® RNA HS Assay Kit (all Invitrogen, Reinach, Switzerland) on the Qubit® Fluorometer 2.0 (Invitrogen, Reinach, Switzerland) according to manufacturer's protocol. Following adjustment of each step, a test NGS run was performed to compare the amount of virus genomes in samples enriched with the established method to non-enriched samples.

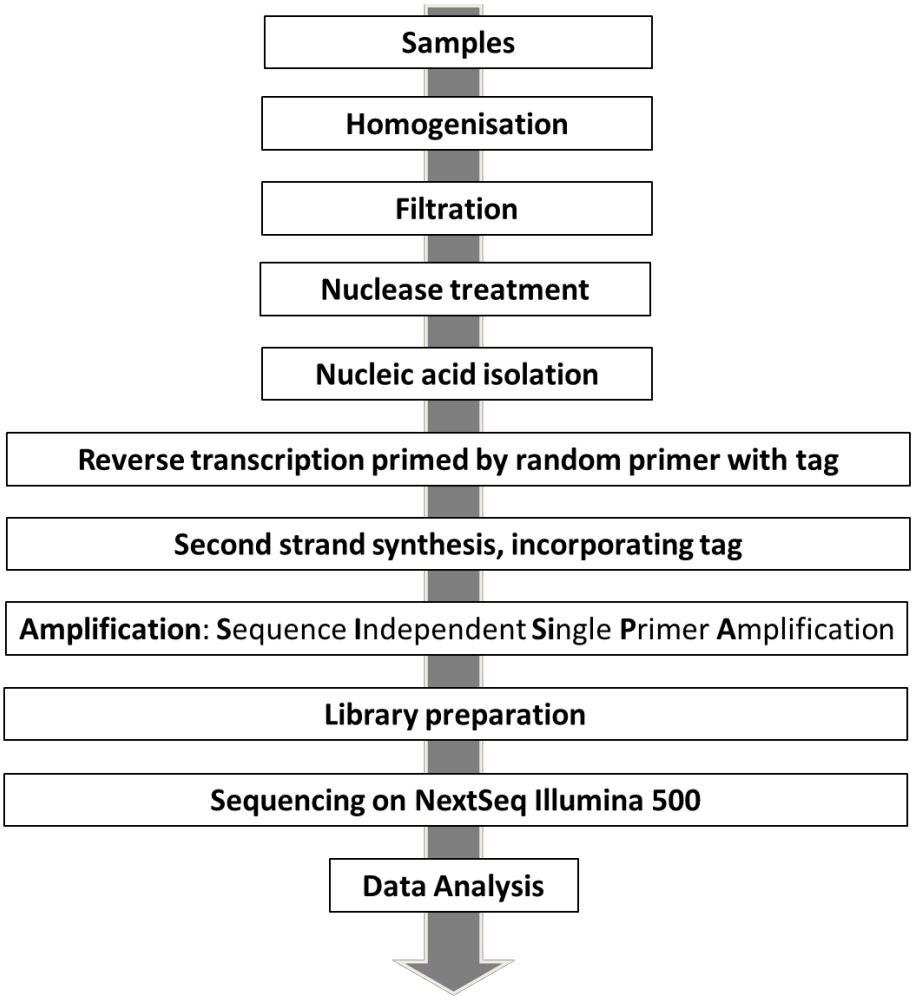

Fig. 1 Schematic representation of the steps included in sample preparation for establishment of the virome protocol.

3.2 Animals, samples, sources of the samples

For technical development of the virome protocol, and to test whether it is applicable in veterinary diagnostics, diverse samples were used. The main group of samples used for testing particular steps of the protocol consisted of swine lungs and feces that had been submitted to the Institute of Veterinary Pathology at the Vetsuisse-Faculty in Zürich. Addition of known cell culture grown viruses was performed to mimic infected material, so-called spiking. Samples were stored at -20°C until usage. Additionally, after establishing the virome protocol using mainly spiked material, it was tested using several naturally infected clinical samples that were sent to the diagnostic unit of the Institute of Virology of the Vetsuisse-Faculty Zürich or used in other research projects. These samples were first tested with routine diagnostics methods,

i.e., qPCR and were confirmed to be positive for specific viruses. All clinical samples included in the study are listed in Table 1.

Number	Species	Material	Diagnosis by specific PCR
1	bison	spleen	BVDV
2	sheep	skin crust	Orf virus (Parapoxvirus ovis)
3	pig	liver	PCV-2
4	pig	feces	PRV-A
5	pig	nasal swab	SIV N1H1
6	human	feces	HEV
7	Pig/Pork	sausage (mortadella di fegato crudo)	HEV

Tab.1. Clinical samples that were tested by specific PCRs and NGS.

3.3 Spiked-in viruses for tests

3.3.1 Bovine viral diarrhea virus (BVDV)

This virus was isolated by Dr. Claudia Bachofen from leukocytes of a BVDV persistently infected animal named Carlito (Stalder et al., 2015). It was passaged 6 times on Madin-Darbey bovine kidney (MDBK) cells, and virus stocks were kept at -80°C.

3.3.2 Bovine Herpesvirus 1 (BoHV-1)

The BoHV-1 field strain used for spiking originates from the last outbreak of IBR in Switzerland in the canton Jura (Blickenstorfer et al., 2010) and was provided by the diagnostic unit of the Institute of Virology. The virus had been passaged 7 times on MDBK cells, and aliquots were kept at -80°C.

3.3.3 Swine influenza A virus (SIV)

The swine influenza virus strain Puerto Rico 8 (PR-8; A/Puerto Rico/8/1934(H1N1)) was originally isolated from a diseased human in Puerto Rico in 1934. It is thought to be at least partially of porcine origin and is related to the pandemic swine flu virus from 2009 (Xu, 2010). It was provided by the diagnostic unit of the Institute of Virology where it serves as positive control for the SIV real-time RT-PCR. It was passaged on Madin-Darbey canine kidney (MDCK) cells.

3.3.4 Porcine Rotavirus A (PRV-A)

OSU is a PRV-A reference strain frequently used in Rotavirus research and -diagnostics (Rushlow et al., 1988). It was provided by Dr. Catherine Eichwald who passaged it on MA104 cells (embryonic rhesus monkey kidney, ATCC® CRL-2378) (Eichwald et al., 2012).

3.3.5 Porcine Parvovirus (PPV)

The porcine parvovirus isolate used was provided by the diagnostic unit of the Institute of Virology where it is used as positive control in the immunofluorescence test. It had been isolated from a diagnostic sample (PCR positive fetal tissue) and passaged 6 times on porcine kidney cells (PK-13).

3.3.6 Transmissible Gastroenteritis virus (TGEV)

The TGEV reference strain "Purdue" was originally isolated at Purdue University, Indiana, USA, and is a widely used TGEV lab strain that is fully sequenced (Zhang et al., 2007). It was provided by the diagnostic unit of the Institute of Virology where it is used as positive control for TGEV real-time RT-PCR.

3.4 Homogenization

Proper homogenization is the first step of preparing biological samples for nucleic acid isolation as it releases the viruses from cells and reduces filter clogging during filtration. Fifty milligrams of feces and tissue (lung) were weighed, added to 250ul of PBS and spiked with BVDV, PRV-A and TGEV. Next, the three samples were divided into two groups. The first group was homogenized with stainless steel beads of 5mm diameter (QIAGEN, Hombrechtikon, Switzerland) and the second was first cut with a sterile scalpel blade on a petri dish into very small pieces (approx. 1mm^2 diameter) and then homogenized as the first group but without steel beads. Additionally, BVDV infected cells were also compared in homogenization with and without beads. Homogenization was done using the TissueLyser II (Qiagen) for 2 minutes at 20Hz. Afterwards, nucleic acid was extracted using peqGold TrifastFL (Peqlab) as described below (see 3.6 Nucleic acid isolation) and quality and quantity measured on the Qubit™ 2.0 Fluorometer (Invitrogen, Reinach, Switzerland). Detection of PRV-A and TGEV RT-qPCR was performed using VetMAX™ Swine Enteric Panel (Applied Biosystems, Switzerland) according to the manual protocol using the following conditions: 10 min at 48°C, 10 min at 95°C, and 45 cycles of 15 sec at 95°C and 1 min at 60°C. Primers and Probes used for BVDV detection are listed in Table 2. For the BVDV real-time RT-PCR, 2µl of RNA was used in a

final volume 20µl using the TaqMan® RNA-to-Ct™ 1-Step Kit (ThermoFisher, Reinach, Switzerland) and the RT-qPCR conditions: 30 min at 48°C, 10 min at 95°C, and 45 cycles of 15 sec at 95°C and 1 min at 60°C.

Primer/Probe	**Sequence BVDV**
Forward (20µM)	5′-CCA TGC CCT TAG TAG GAC TAG C-3′
Reverse (20µM)	5′-TGA CGA CTA CCC TGT ACT CAG G-3′
Probe (5µM)	5′-FAM AAC AGT GGT GAG TTC GT MGB-3′

Primer/Probe	**Sequence BoHV-1**
Forward (10µM)	5′-TGT GGA CCT AAA CCT CAC GGT-3′
Reverse (10µM)	5′-GTA GTC GAG CAG ACC CGT GTC-3′
Probe (10µM)	5′-FAM-AGG ACC GCG AGT TCT TGC CGC-TAMRA-3′

Tab. 2. Primer and probe sequences for BVDV and BoHV-1 real-time PCR.

3.5 Filtration

Filtration is a widely used method to remove bacterial and host cells. In combination with centrifugation, it provides an effective way to enrich for viral particles. For filtration, the following pore sizes were tested: 0.45µm (Puradisc, 13mm, Whatman GE Healthcare) and 0.8µm syringe filters (Vivaclear Mini 0.8µm Pes, Sartorius). After homogenization and spiking BVDV and BoHV-1, porcine fecal samples were centrifuged at 13000rpm (Biofuge Fresco, Heraeus) for 5 minutes, supernatants were aspirated with 0.7x32mm needles and pushed through the filters placed on a 2ml syringe. For detection of BVDV, RT-qPCR was performed as described above. For BoHV-1 DNA detection, 10µl of DNA was added to a final volume of 25µl PCR mix containing BoHV-1 specific primers (Tab. 2.) and TaqMan™ Universal PCR Master Mix using the following qPCR conditions: 2 min at 50°C, 10 min at 95°C, and 45 cycles of 15 sec at 95°C and 1 min at 60°C.

3.6 Nuclease treatment

To degrade exogenous nucleic acid without affecting nucleic acid protected by the viral capsid, three different nucleases were tested: Benzonase Nuclease (Merck, Germany), Ribonuclease A from bovine pancreas (Merck, Germany) and Micrococcal nuclease (New England Biolabs, UK). Benzonase is an endonuclease of 30 kDa subunits from gram-negative bacteria, *Serratia marcescens* and degrades all forms of DNA and RNA. It is assumed to completely digest nucleic acids to 5′- monophosphate terminated oligonucleotides of 3 to 5 bases in length.

Ribonuclease A is an endoribonuclease that starts at the 3'OH phosphate of a pyrimidine nucleotide in a single stranded RNA. Micrococcal nuclease is a nonspecific endo-exonuclease derived from *Staphylococcus aureus* and digests double-stranded, single-stranded, circular and linear RNA and DNA. Lung samples spiked with BVDV and BoHV-1 were divided into 7 groups that were treated either with one or a combination of two or three nucleases. The same amount (1µl) of each nuclease was used in all 7 combinations with 132.5µl filtrated and homogenized sample and incubated in a thermal cycler for 15 min at 45°C followed by 1 hour at 37°C. After incubation, total RNA and DNA was measured on the Qubit fluorometer and RT-qPCR for BVDV and BoHV-1 was performed as described above. Additionally, after selection of nucleases, a spleen sample spiked with BVDV and BoHV-1 was compared with and without nucleases. As an additional way to monitor the efficiency of the nucleases in degrading host nucleic acid, this sample was also tested for the reduction of porcine mitochondrial ribosomal RNA and its coding region (p12s).

3.7 Nucleic acid isolation

To isolate nucleic acid, two methods were compared: the guanidine isothiocyanate based peqGold TriFastFL (Peqlab), and the silica column based Qiamp Viral RNA mini kit (Qiagen GmbH, Germany). Both methods were performed according to the manufacturers manuals with exception of: using aqueous phase containing RNA and interphase containing DNA simultaneously for peqGold TriFastFl RNA isolation, omitting the carrier RNA in the viral RNA kit and adding 1% of β-mercaptoethanol (Bio-rad) to the AVL buffer in order to efficiently inactivate the nucleases of the preceding step (3.5). Nucleic acid of two spleens and one lung sample spiked with BVDV and BoHV-1 was extracted by both methods and RT-qPCR for BVDV and BoHV-1, as well as nucleic acid concentration measurements were performed as described above.

3.8 Reverse transcription and second strand synthesis

To sequence both, RNA and DNA viruses simultaneously, RNA needs to be reverse transcribed into cDNA. For cDNA synthesis 2.5 µM of a random primer with a known 20nt-tag sequence, SISPA-N (GCT GGA GCT CTG CAG TCA TCN NNN NN), was added to the RNA. Two cDNA synthesis kits were compared: RevertAid First Strand H minus cDNA Synthesis Kit (ThermoFisher, Reinach, Switzerland) and SuperScript IV Reverse Transcriptase (Thermo Scientific) on three BVDV spiked-in spleen samples. Both kits were used according to manufacturer's protocols. To degrade remaining RNA after cDNA synthesis, 1µl of RNase H

(New England Biolabs, BioConept, Allschwil, Switzerland) was added up to 100 µl reaction volume and incubated at 37°C for 20 minutes. For the second strand synthesis, to 45.5 µl of first strand DNA a premix of primer SISPA-N (GCT GGA GCT CTG CAG TCA TCN NNN NN), 10x Klenow Buffer and dNTP (10mM) was added and denatured by incubating at 95°C for 1 min and then cooled on ice. After the denaturation step, Klenow Fragment 3'→5' exo- (ThermoFisher, Reinach, Switzerland) was added to synthesize the second strand and incubated according to manufacturer's instruction. After incubation, samples were purified using PureLink® PCR Micro Kit (Invitrogen-ThermoFisher, Reinach, Switzerland) according to the manual and eluted in 12 µl elution buffer.

3.9 Random amplification

Since filtration and nuclease treatment dramatically reduce the amount of total nucleic acids, the remaining DNA has to be amplified to obtain sufficient material for library preparation and sequencing. Therefore, sequence-independent single primer amplification (SISPA) was used to amplify DNA isolated from lung samples spiked with BoHV-1 and BVDV (Fig. 2.). For the unspecific PCR reaction, we used HotStarTaq DNA polymerase (QIAGEN, Hombrechtikon, Switzerland) and the primer SISPA (GTT GGA GCT CTG CAG TCA TC) that binds to the complementary sequence of the tag introduced with the SISPA-N (GTT GGA GCT CTG CAG TCA TC NNN NNN) primer in first and second strand synthesis. The number of amplification cycles was set to 10, 15, 18, 20 or 25 in order to define the lowest possible but highest necessary number of cycles. The amplified products were purified using the QIAquick PCR Purification kit (Qiagen) according to manufacturer's protocol and eluted in 30 µl elution buffer. The amount of DNA was measured using the Qubit fluorometer and RT-qPCR for BVDV and qPCR for BoHV-1 performed as described above.

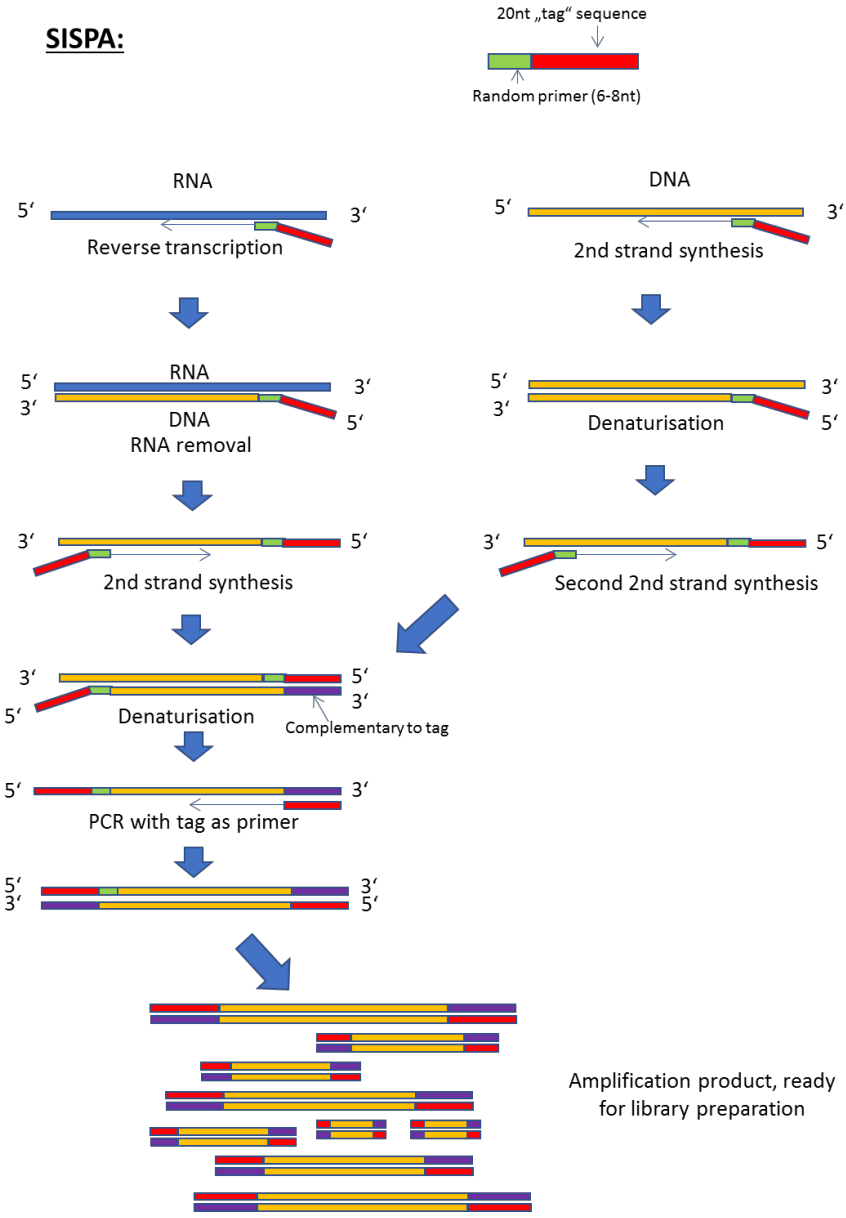

Fig.2. Schematic representation of sequence-independent single-primer amplification (SISPA) used for random amplification.

3.10 Library preparation and NGS setup

For performing NGS, libraries specific for each NGS platform are required. Library preparation includes fragmentation of DNA, which is followed by ligation of specific adaptors to the DNA fragments. In the Illumina system the adapters are used to fix the DNA fragments to the so-called flow cell where sequencing occurs. The adapters also contain primer sites for clonal amplification of the fragments (bridge PCR) prior to sequencing and for the sequencing reaction itself. In a test run, two commercial kits for library preparation were compared: NEBNext®

Ultra™ I DNA Library Prep Kit and NEBNext® Ultra™ II DNA Library Prep Kit (both New England Biolabs, BioConept, Allschwil, Switzerland). Three porcine fecal samples were spiked with 6 viruses and enriched for virus DNA as described before. The first step of library preparation encompassed fragmentation of the DNA to the desired length of 500bp by acoustic shearing on the E220 focused-ultrasonicator (Covaris, Massachusetts, USA). The length of the DNA was checked on an Agilent 2200 TapeStation using D1000 ScreenTape assay. The libraries were prepared from 50ng DNA input for the Ultra I kit and from 50 ng and 1 ng DNA input for the Ultra II kit according to instructions in the manual. Six and eight PCR cycles were performed for 50ng and 1ng DNA input, respectively. Cleanup of adaptor-ligated DNA was performed without size selection by AMPure XP Beads (Beckman Coulter). For barcoding, NEBNext Multiplex Oligos (New England BioLab) index primers were used. The length and molarity of the sequencing libraries were checked on the Agilent 2200 TapeStation (Agilent, USA) and by qPCR on a Lightcycler 96 (Roche, Switzerland). For qPCR quantification, samples as well as standard Illumina PhIX Control, were diluted in 0.1% Tween 20. KAPA SYBR FAST qPCR Kit (Roche) was used to perform analysis. Finally, the libraries were diluted to equal molarity according to the qPCR results and pooled. The starting library concentration for the Illumina NextSeq 500 was 0.5nM. A paired-end NGS run of 2 × 150 nucleotide read length, using the high-output flow cell was performed at the Functional Genomics Center Zurich using the Illumina NextSeq 500 system.

3.11 Testing of the virome protocol using spiked samples

To check the effect of the enrichment for virus particles on the sequencing results, i.e., recovery of viruses, 3 types of samples: feces, lung tissue and nasal swab were spiked with 6 known viruses. To verify the ability of the method to detect viruses of different families and with different genome sizes, the following viruses were spiked into the samples: bovine herpes virus 1 (BoHV-1, dsDNA), porcine parvovirus (PPV, ssDNA), rotavirus A (RV-A, segmented dsRNA), bovine diarrhea virus 1 (BVD-1, +ssRNA) transmissible gastroenteritis virus (TGEV, +ssRNA) and Influenza A virus (N1H1, segmented -ssRNA). All viruses were from stocks of the Institute of Virology, Vetsuisse-Faculty, Zürich. Afterwards, the 12 spiked samples were divided into four groups, according to the preparation methods applied:

1. Samples prepared with enrichment methods such as filtration and nuclease treatment and 18 cycles of DNA amplification;

2. Samples prepared with enrichment methods such as filtration and nuclease treatment but without amplification;

3. Samples prepared without enrichment methods but with DNA amplification;

4. Samples prepared without enrichment method and amplification.

3.12 Testing the virome protocol with clinical samples

In the second sequencing run, clinical samples from the Vetsuisse Faculty clinics that were tested in the Diagnostic department at the Institute of Virology and showed positive results in specific qPCRs were prepared and sequenced (see Tab.1.). Additionally, a not diagnosed brain sample of a 4 months old pig with neurological symptoms was also added. In histopathological examination done at the Institute of Veterinary Pathology, Vetsuisse-Faculty, Zürich, multifocal lymphocyte infiltration in grey and white matter was observed and suspicion was Polioencephalomyelitis enzootica suum caused by porcine Teschovirus-1 (Enterovirus, Picornaviridae).

3.13 Testing the sensitivity of the protocol

To test the sensitivity of virus detection by NGS, a pig nasal swab sample was split into five tubes and spiked with decreasing concentrations of Influenza A virus (strain PR8) from a previously prepared 10-fold dilution series that showed the Ct values: 25.8; 29.4; 33.5; 36.5; 38.1 in a specific RT-qPCR. Thereafter samples were prepared according to the virome protocol.

3.14 Full-length sequencing of Hepatitis E virus (HEV)

A clinical sample from a hospital in Lugano and from the cantonal laboratory in Bellinzona, Switzerland, were sequenced: stool sample of a 78-year-old male with acute hepatitis and dry cured sausage produced with raw liver from pigs (mortadella di fegato cruda). There was a suspicion that the raw cured sausage was the source for the hepatitis E infection. Both samples were confirmed as positive for Hepatitis E virus (HEV) by RT-qPCR using the commercial kit (hepatitis@ceeramTools®, bioMérieux, Geneva, Switzerland) according to the manufacturer instructions. Both samples were prepared according to virome protocol.

3.15 Whole genome sequencing and Swiss SIV surveillance program

To investigate if our protocol is also applicable for whole genome sequencing and may be used for the Swiss swine influenza virus surveillance program, we compared it to the protocol used

up to now that is based on targeted amplicon sequencing. For that reason, the same influenza positive (CT 25) pig nasal swab sample was prepared with the PathAmp™ FluA Reagents (ThermoFisher scientific) according to the producer's manual and with our virome protocol. For the virome protocol, samples were prepared as follows: (I) standard protocol but with previous concentration of the sample in an Amicon® Ultra 15 mL Centrifugal Filter (100k Molecular Weight Cutoff- MWCO) (Merck) from 1200µl to 200µl, (II) standard virome protocol using 134ul of the swab medium, and (III) double amount of sample was used for nucleic acid isolation (2x 134µl of supernatant). Afterwards, libraries were prepared from all 3 samples and sequenced on MiSeq (Illumina, USA).

3.16 Data analysis

In a first step, all generated raw data were quality checked with FastQCapp available in the workflow management system SUSHI (https://fgcz-sushi.uzh.ch/), provided by the Functional Genomics Center Zürich (Hatakeyama et al. 2016). Thereafter, reads were aligned to the latest available NCBI virus reference sequences database (ftp://ftp.ncbi.nlm.nih.gov/refseq/release/viral/) using the metagenomic pipeline of the SeqMan NGen software v14 (DNAStar, Lasergene, Madison, WI, USA). The resulting alignments were visualized and analyzed using the SeqMan PRO software v14 (DNAStar, Lasergene, Madison, WI, USA).

4. RESULTS

4.1 Homogenization

Homogenization is a step that allows preparing a sample for further analysis. In case of viruses that are located intracellularly, it is necessary to lyse or break the cells to release their content. However, extensive homogenization may damage viral particles and release host genomic DNA and RNA. Therefore, two feces and one lung sample that were spiked with BVDV, PRV-A and TGEV, were homogenized in two ways: with 5mm metal steel beads and by just cutting into small pieces and homogenizing without bead. Both samples were homogenized in the TissueLyser. Results from RT-qPCR showed that the CT values of RNA and DNA virus were generally very similar for both homogenization methods (Fig.2). In most cases, the CT-values were slightly lower if homogenization was performed with the beads. However, for the lung and feces 2 the TGEV CT was higher with the beads. Since spiked-in viruses are not intracellularly located, we tested also BVDV infected cell cultures. The CT value was 22.4 with the beads and 22.5 without the beads (data not shown). Measurements of total DNA and RNA showed that there was slightly less DNA (Fig.3) and clearly less RNA in samples homogenized without the metal steel beads.

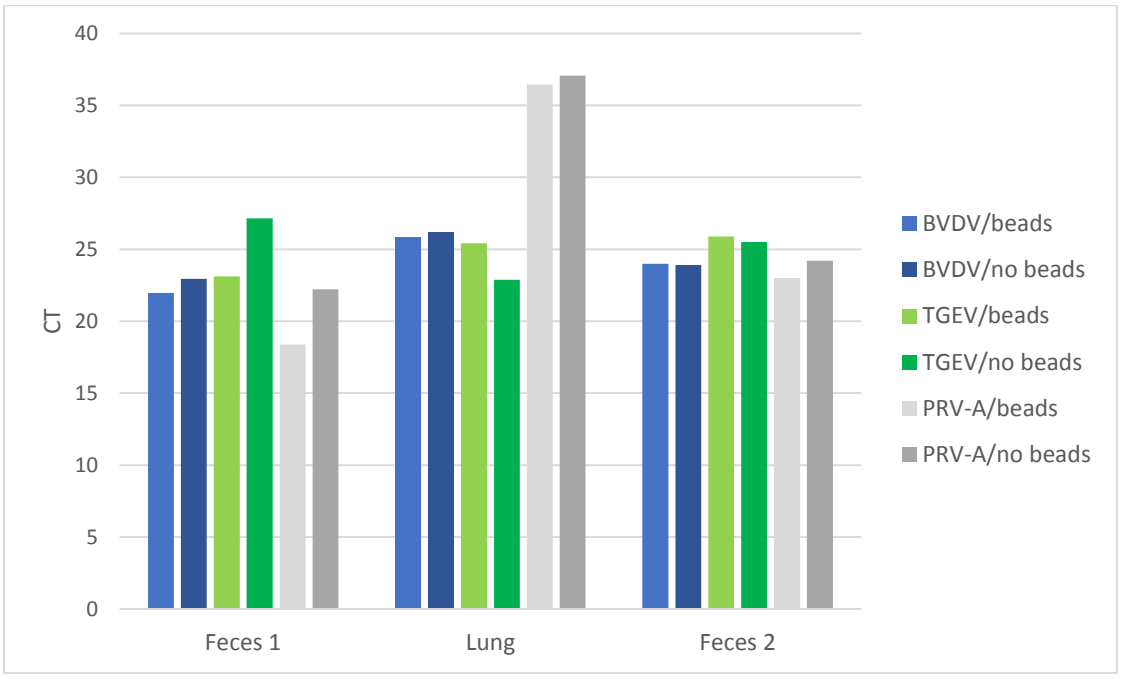

Fig.2 Ct values of spiked-in Bovine viral diarrhea virus (BVDV), transmissible gastroenteritis coronavirus (TGEV) and porcine rotavirus A (PRV-A) for two different feces samples and one lung sample homogenized with/without 5mm steel beads.

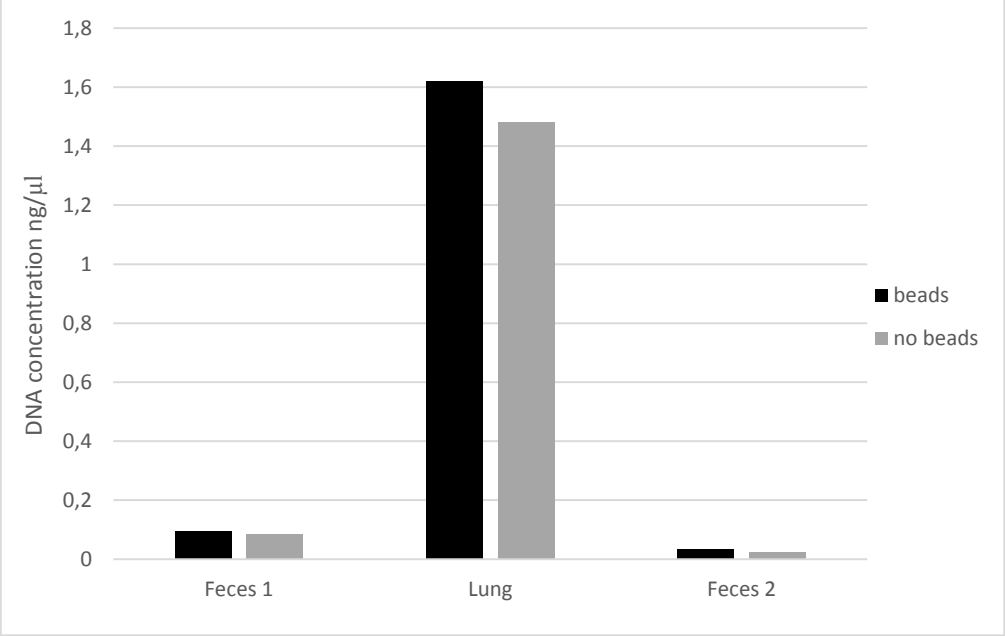

Fig. 3 Total DNA concentration (ng/ µl) of two feces and one lung sample homogenized with or without 5mm steel beads.

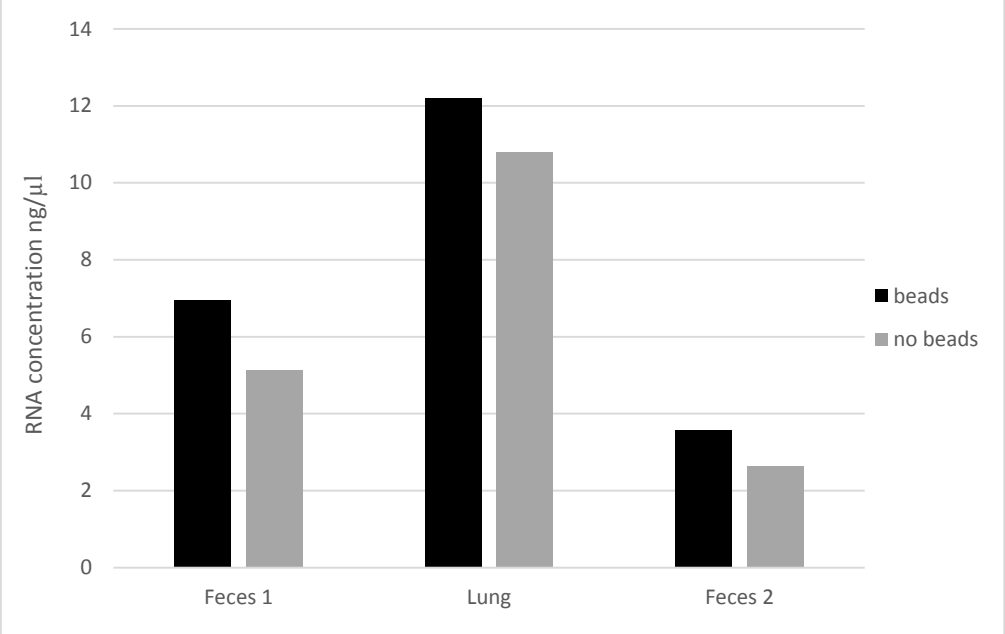

Fig. 4 Total RNA concentration (ng/µl) of two feces and lung sample homogenized with or without 5mm steel beads.

4.2 Filtration

Filtration is applied to remove cell debris and bacteria without losing viruses. The Ct values of BVDV and BoHV-1 in the RT-qPCR were nearly on the same level (Fig.5). However, the total

concentration of DNA and RNA was slightly lower for sample treated with 0.45µm syringe filter than using 0.8µm syringe filter (Fig. 6). Both filters that were used appeared to provide the same amount of viral genome. Nevertheless, the 0.45µm syringe filter ought to provide greater reduction of bacteria and cells.

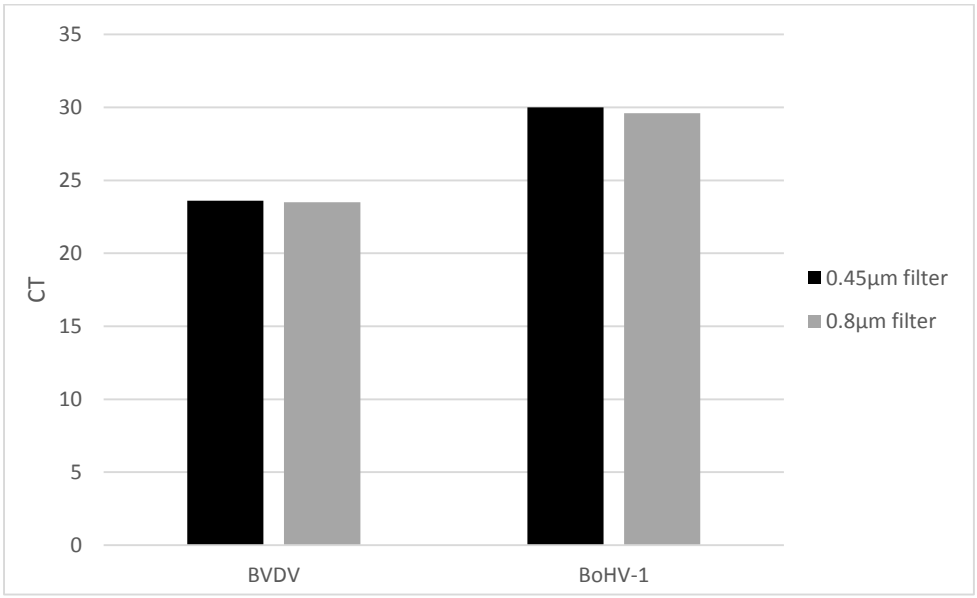

Fig. 5 Ct values of spiked-in BVDV and BoHV-1 of the fecal sample prepared with two different sized syringe filters: 0.45µm and 0.8µm.

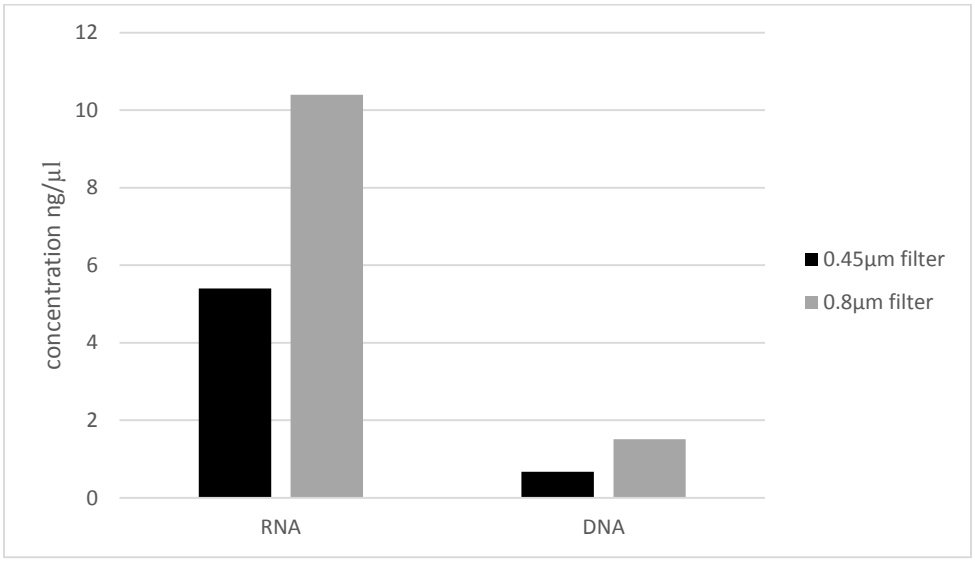

Fig. 6 Total RNA and DNA concentration of the fecal sample prepared with two different sized syringe filters: 0.45µm and 0.8µm.

4.3 Nuclease treatment

The aim of the nuclease treatment was to remove as much non-viral nucleic acids as possible while leaving viral genomes unaffected. Therefore, lung samples were spiked with RNA virus (BVDV) and DNA virus (BoHV), homogenized and treated with different single and combined nucleases. There was no significant difference in the amount of BVD RNA or BoHV-1 DNA detected in all seven groups treated with various nucleases (Fig. 7 and 8). Slightly lower Ct value and therefore more virus was obtained from the sample treated only with Benzonase for BVDV and with RNase A for BoHV-1. However, measurements of the concentrations of total DNA and RNA (Fig. 9) showed that Benzonase alone provided the lowest reduction of total RNA. The greatest RNA and DNA reduction was observed from the combined treatment with RNase A and Micrococcal nuclease. In addition, the price of those two nucleases together is significantly lower than of Benzonase. The total DNA and RNA concentrations in the untreated sample was clearly the highest (Fig 9.). RNA was even above detection limit, which is 100ng/µl for the Qubit RNA assay. While nuclease treatment in a sample spiked with BVDV and BoHV-1 reduced viral RNA and DNA only slightly, the host genomic nucleic acids as measured by the housekeeping gene p12S was substantially reduced (Fig. 10).

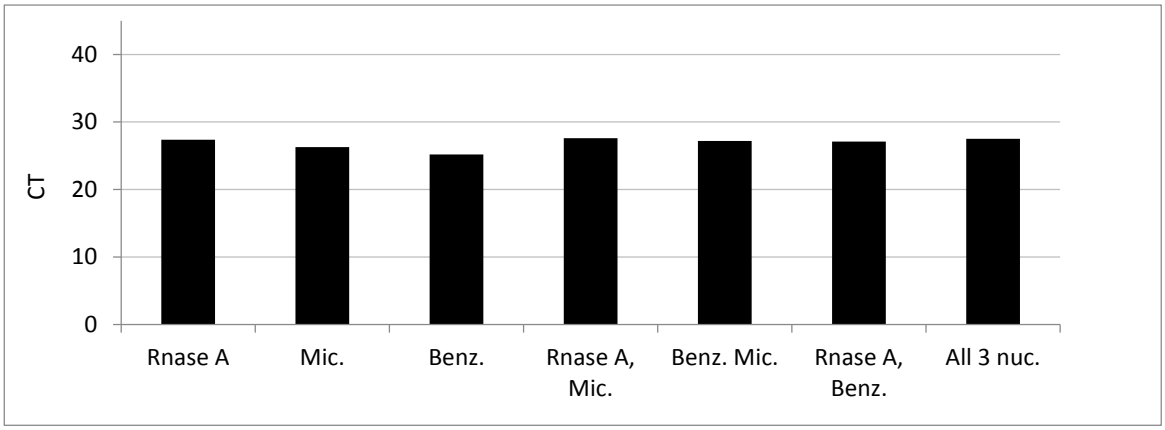

Fig. 7 BVDV CT values of 7 samples prepared with combinations of three different nucleases. Benzonase (Benz.), Ribonuclease A (Rnase A), Micrococcal nuclease (Mic.).

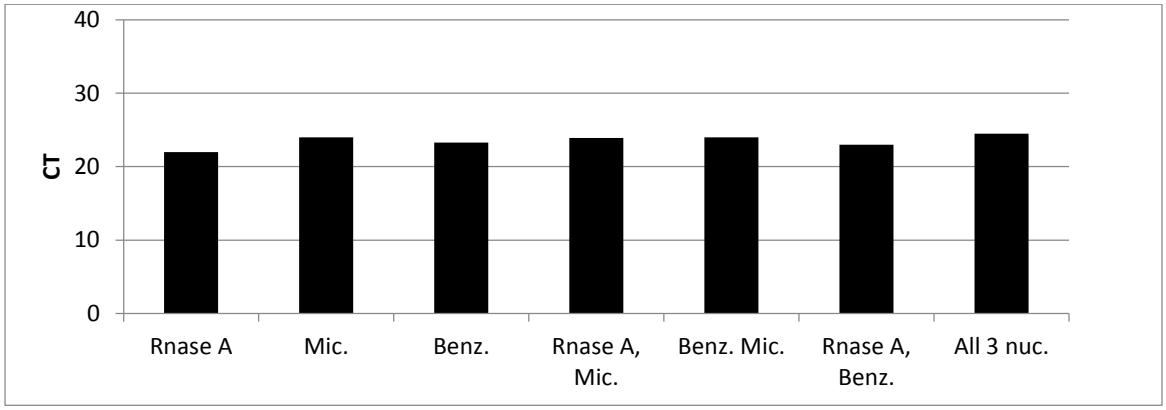

Fig. 8 BoHV-1 CT values of 7 samples prepared with combinations of three different nucleases. Benzonase (Benz.), Ribonuclease A (Rnase A), Micrococcal nuclease (Mic.).

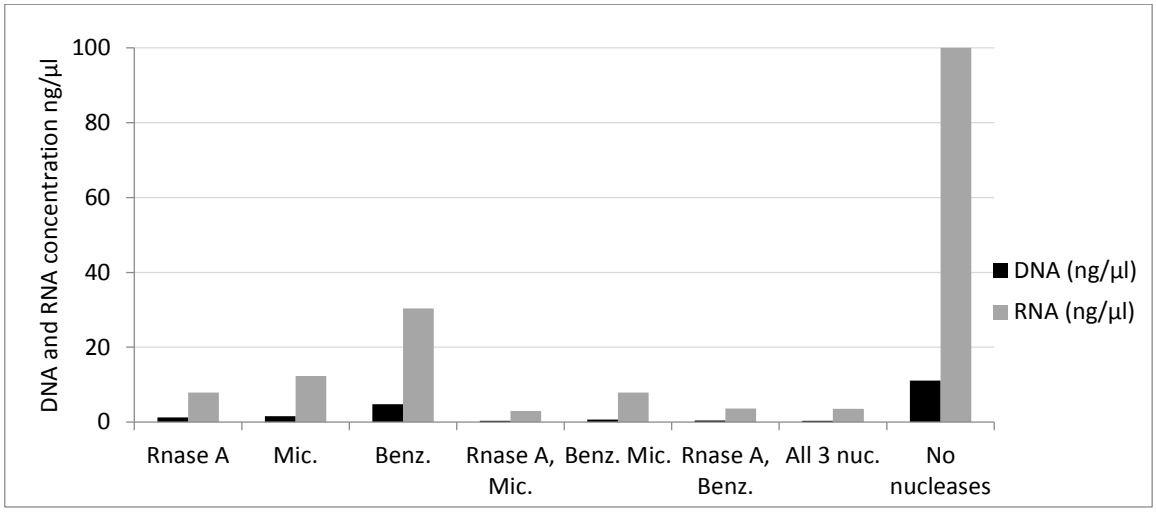

Fig. 9 Total DNA and RNA concentration of lung samples trated with combination of nucleases. Benzonase Nuclease, Ribonuclease A, Micrococcal nuclease. RNA concentration for sample without nucleases is above detection limit (100ng/µl).

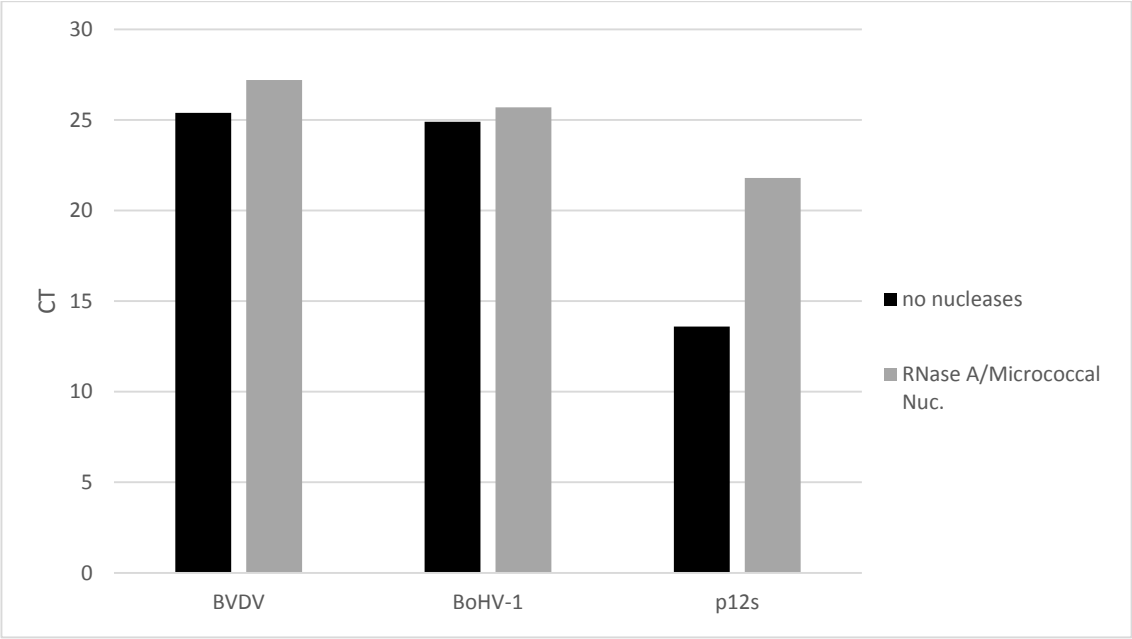

Fig.10 Ct values of BVDV, BoHV-1 and p12s in a spleen sample treated with RNase A and Micrococcal Nucleases and without any nucleases treatment.

4.4 Nucleic acid isolation

Both nucleic acid isolation methods that were used for comparison (Phenol-chloroform based and silica columns) can be used to isolate DNA and RNA. However, we wanted to know which results in highest viral RNA and DNA and lowest total RNA and DNA concentrations. Therefore, we compared them on two different lungs and one spleen sample spiked with DNA virus (BoHV-1) and RNA virus (BVDV). The results showed that peqGold TrifastFL provides lower CT values hence more viral RNA for BVDV than the viral RNA mini kit. On the contrary, BoHV-1 was better detected by the viral RNA mini kit than the TriFastFL (Fig 11.). The total DNA concentration in the sample isolated with the Viral RNA mini kit was clearly lower than with peqGold TrifastFL. Due to the treatment with nucleases prior to nucleic acid isolation, total RNA was below the detection limit which was 250pg/µl for the Qubit RNA assay used.

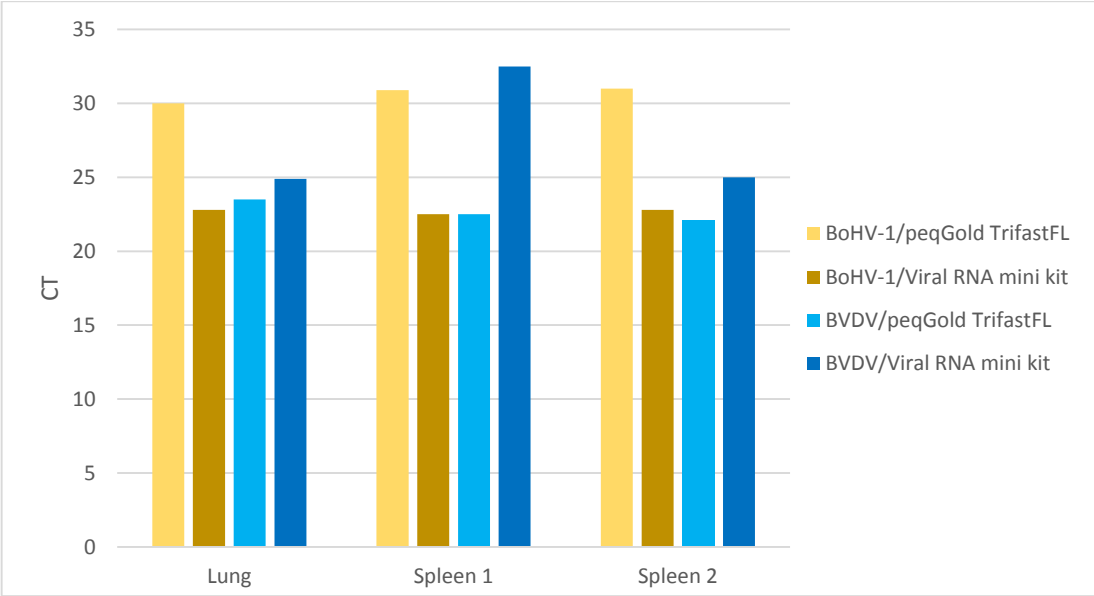

Fig. 11 Ct values of two different spleen and one lung sample spiked with BoHV-1 and BVDV. RNA and DNA was extracted using peqGold TrifastFL and Viral RNA mini kit.

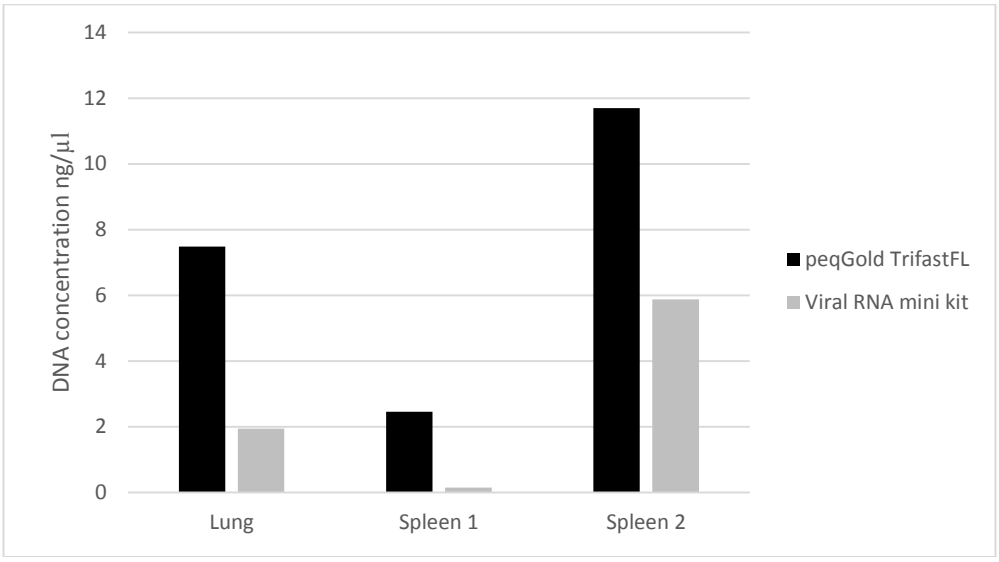

Fig. 12 Concentration of total DNA isolated from the spleen and lung samples spiked with BoHV-1 and BVDV, extracted with peqGold TrifastFL and viral RNA mini kit. Total RNA was in all samples below the detection limit of the test.

4.5 Reverse transcription and second strand synthesis

For the reverse transcription, two commercial kits were tested: RevertAid First Strand H minus cDNA Synthesis kit and SuperScript IV Reverse Transcriptase. Both kits transcribed RNA of BVDV into cDNA. However, the results from qPCR showed that from the same amount of starting RNA, the RevertAid kit resulted in more cDNA (Fig 13.)

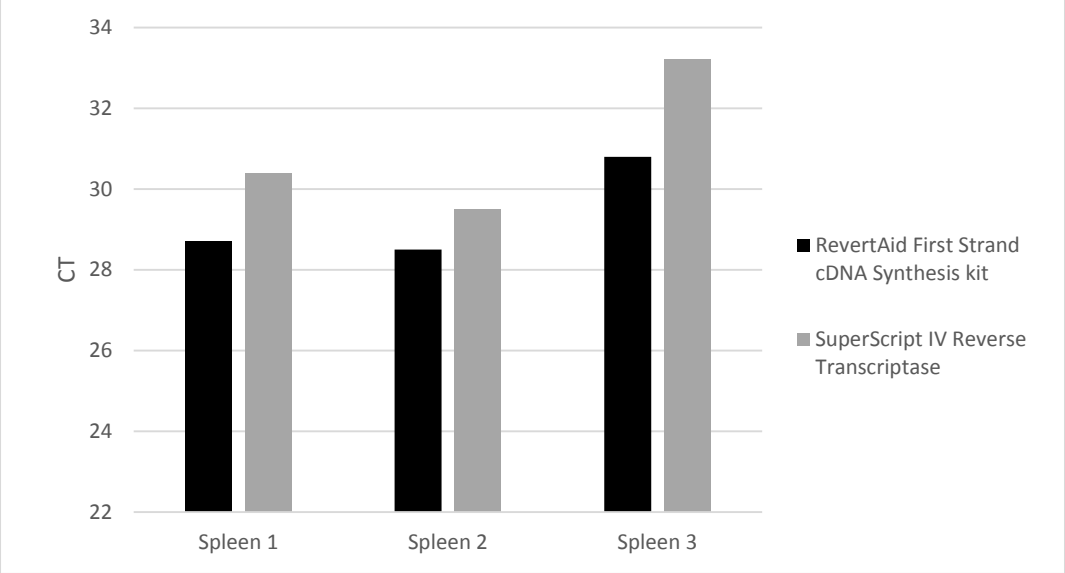

Fig. 13 Ct value of DNA transcribed from BVDV RNA with two kits: RevertAid First Strand cDNA synthesis Kit and SuperScript IV Reverse Transcriptase.

4.6 Random amplification (SISPA)

Since the enrichment method causes significant decrease in the total amount of nucleic acids, amplification is required to provide sufficient amount of material for subsequent library preparation and sequencing. From double-stranded cDNA of lung samples spiked with BoHV-1 and BVDV, sequence-independent single primer amplification (SISPA) was performed. Different numbers of cycles were used for SISPA and the products tested in qPCR and total DNA measurement to determine the lowest possible cycle number resulting in sufficient DNA for downstream processing, i.e., 10, 15, 18, 20, 25 cycles. In the first experiment, 10, 15, 20, 25 cycles were compared. Based on the results, 18 cycles were added for the second experiment using the same starting material. In the last experiments 10 cycles were excluded. This resulted in five repetitions for cycles 15, 20 and 25, three repetitions for 18 cycles and two for 10 cycles. The results from qPCR demonstrated that, as expected, the more amplification cycles were used, the more viral cDNA was observed for BVDV (Fig. 14) as well as BoHV-1 (Fig.15). However, using too many cycles in random amplification causes over amplification primarily of background DNA. As demonstrated in Figure 16, 20 and 25 cycles led to a massive increase of total DNA while the increase of the viral qPCR signal was only moderate.

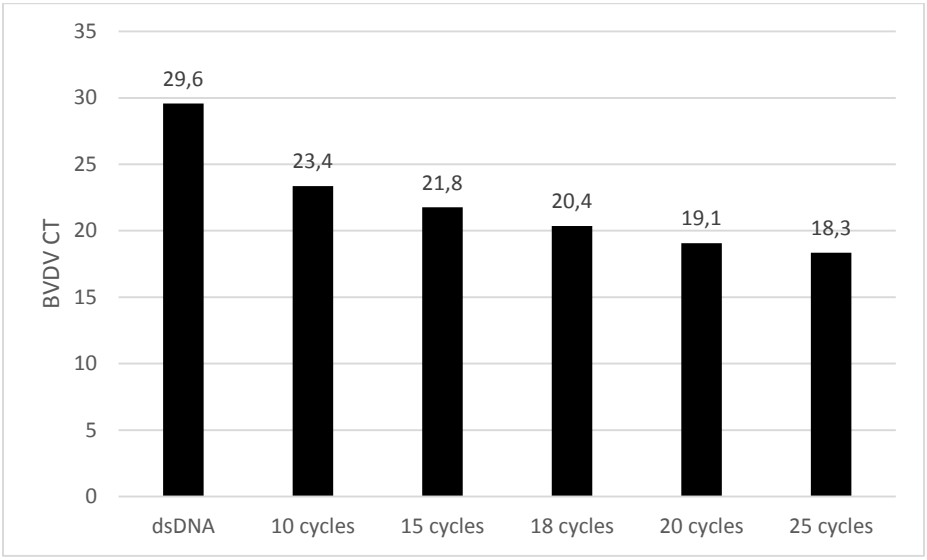

Fig. 14 Mean CT values of BVDV RT-qPCR from double-stranded DNA amplified in PCR run with following numbers of PCR cycles: 25,20,18,15,10. For dsDNA amplification was omitted.

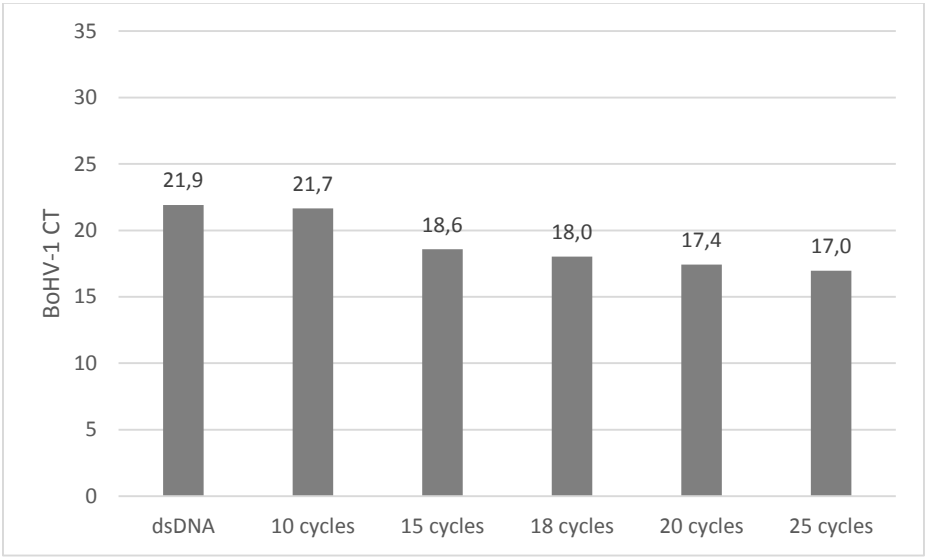

Fig. 15 Mean CT values of BoHV-1 RT-qPCR from double-stranded DNA amplified in PCR run with following numbers of PCR cycles: 25,20,18,15,10. For dsDNA amplification was omitted.

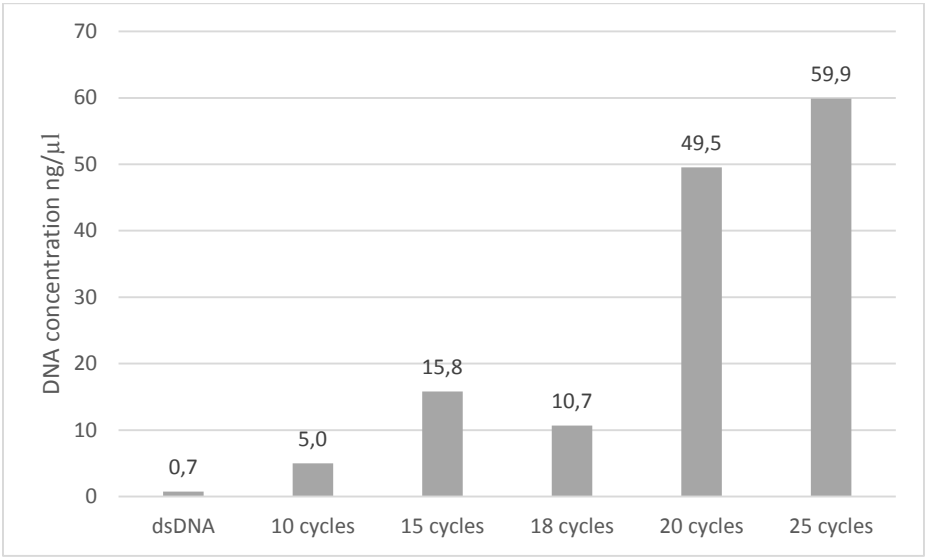

Fig. 16 Mean DNA concentration (ng/µl) from double-stranded DNA amplified in PCR run with following numbers of PCR cycles: 25,20,18,15,10. For dsDNA amplification was omitted.

4.7 Library preparation kits

Two different library preparation kits and different amounts of input DNA were used to prepare libraries from SISPA products of same porcine feces spiked with six different viruses. Using NEBNext Ultra I kit with 50ng DNA input resulted in the highest total number of sequenced reads. This kit enables to prepare libraries from 5ng-1µg of DNA. NEBNext Ultra II kit is a more recent version of Ultra I and enables to prepare libraries with as little input as 500pg of DNA, applying lower numbers of PCR cycles. Interestingly, using Ultra II kit, we obtained more total reads with 1ng input than with 50ng (Fig. 17).

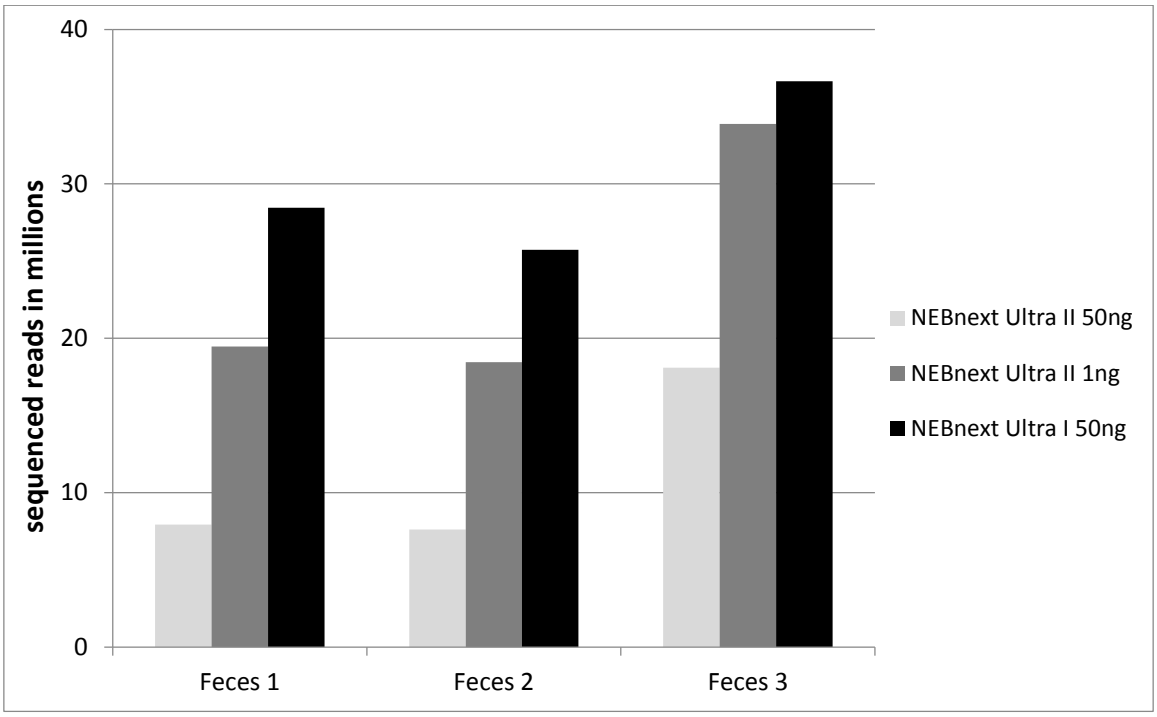

Fig. 17 Total amount of sequencing reads per sample of 3 fecal samples using different library kits and concentration of DNA: NEBNext Ultra II kit 50ng and 1ng DNA input and NEBNext Ultra I 50ng DNA input.

The sequenced reads were assembled to the NCBI viral reference sequence database to compare recovery of the six spiked-in viral genomes with the different library preparations. The results are shown in Fig.18. The NEBNext Ultra II kit with 1ng input, even though providing less total reads than with 50 ng input, resulted in the highest assembly rate.

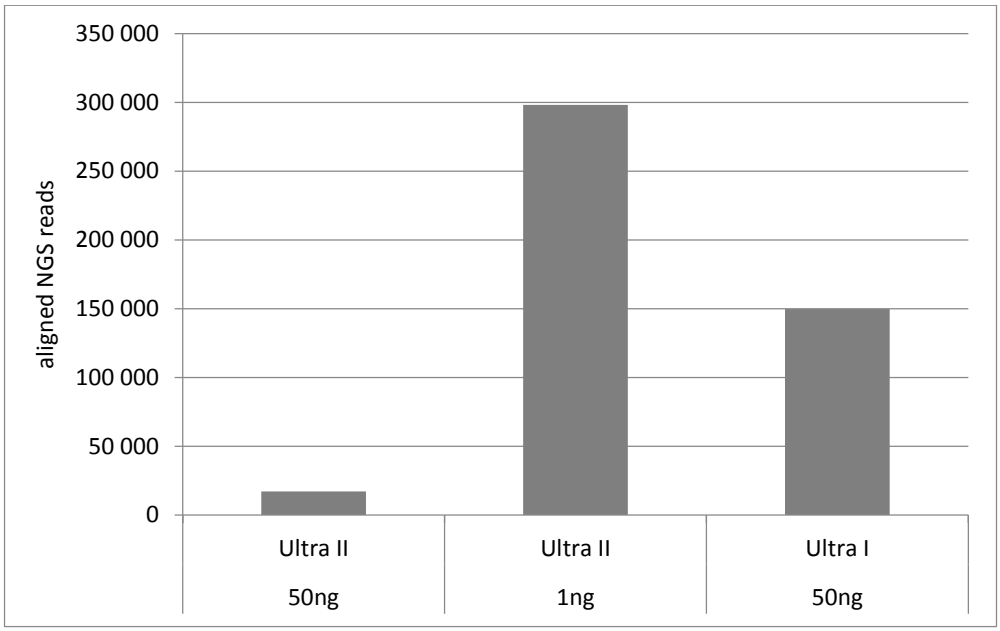

Fig. 18 Mean of total number of reads that were assembled to the reference sequences database of 3 fecal samples using different library kits and concentration of DNA: NEBNext Ultra II kit 50ng and 1ng DNA input and NEBNext Ultra I 50ng DNA input (n=3).

4.8 Testing the virome protocol

Based on the results above, four different sample preparation methods were selected and compared in a first sequencing run to determine the best protocol for virus detection. The data obtained from sequencing three different materials (feces, nasal swab, lung tissue) spiked with six different viruses showed that the samples prepared with the virus enrichment step but without amplification and samples without pre-treatment resulted in the highest number of total reads (Fig.19). For the samples where enrichment and amplification were applied and samples where just the amplification was performed, similar numbers of total reads were achieved (Fig.19). The reads were assembled to a viral database consisting solely of the genomes of the six viruses that were spiked into the samples. In contrast to the total number of reads, the highest number of spiked-in viral reads was observed in the enriched and amplified sample (Fig. 20). The results in Fig. 21 show that not only the absolute number of viral reads but also the relative number, expressed in percentage of total reads, was highest with the full virome protocol. There were 67% matched sequences for nasal swab, 3% for feces and 27% for lung tissue. Furthermore, all spiked-in viruses were recovered, except for the samples without any pre-treatment, where swine influenza A virus was missed in the fecal sample (data not shown).

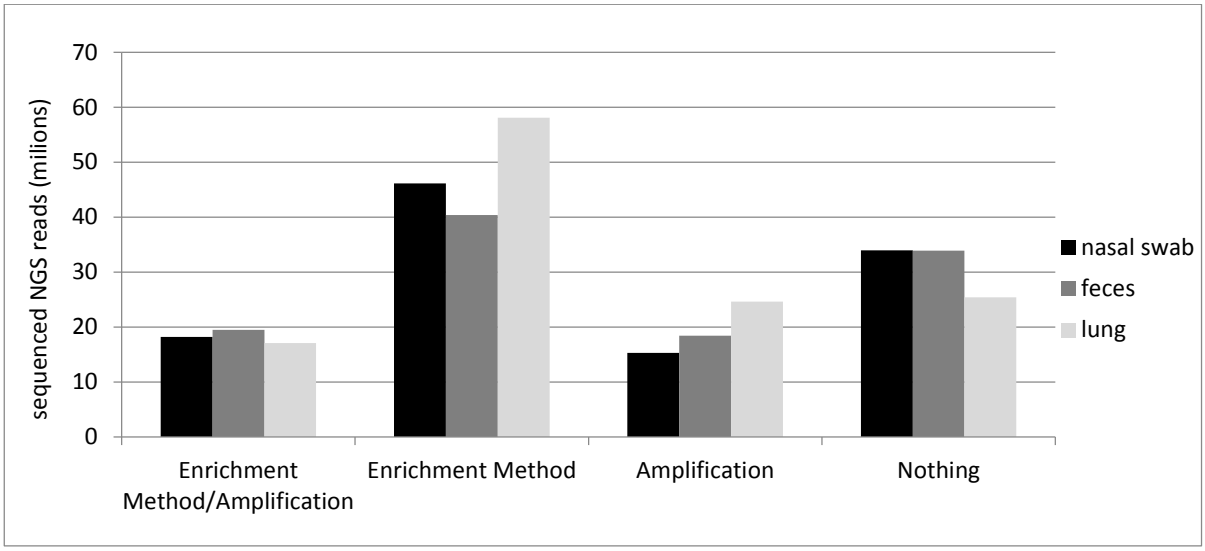

Fig. 19 Total amount of sequenced reads obtained from nasal swab, feces and lung. Samples prepared in 4 different variants of virome protocol. 1. Enrichment method (homogenization, filtration and nuclease treatment) and 18 cycles of random amplification, 2. Enrichment method (homogenization, filtration and nuclease treatment), 3. 18 cycles of amplification. 4. Enrichment method and amplification was omitted.

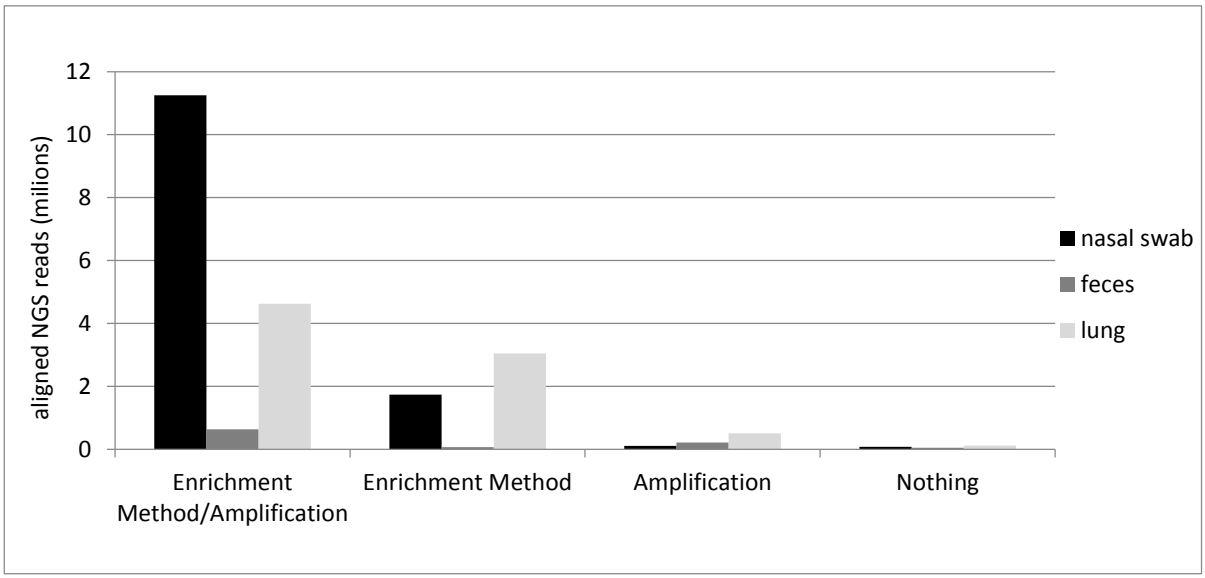

Fig. 20 Numbers of reads from nasal swab, feces and lung matching to the genomes of the six spiked-in viruses.

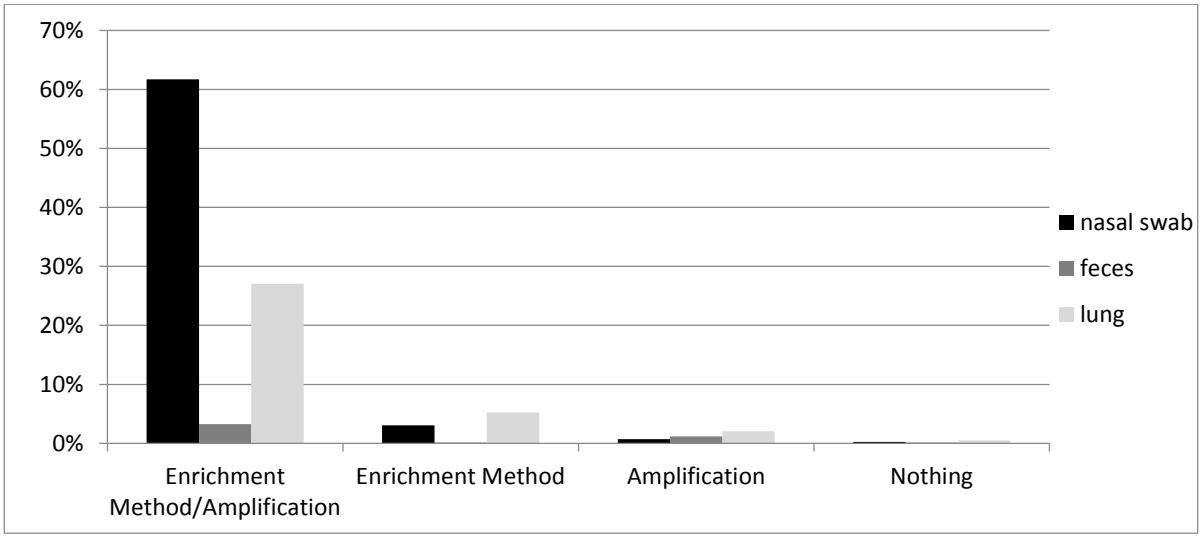

Fig. 21 Viral reads matching to the spiked-in viruses in percentage of the number of total reads

Based on these results, it was decided that the most promising protocol for virome analysis of tissue, feces and nasal swab was the combination of enrichment and amplification. Thus, the final protocol consists of the homogenization with scalpel blade and TissueLyser for 2 min. at 20Hz, centrifugation at 14000rpm for 5 minutes, 0.45 µm syringe filter filtration, treatment with RNase A and Micrococcal Nuclease, nucleic acid extraction with the Viral RNA mini kit, reverse transcription with the RevertAid kit and 18 cycles of random amplification (SISPA). The libraries are prepared from 1ng of DNA by the NEBNext Ultra II kit (Fig.24).

4.9 Clinical samples

Thus far, due to accessibility and repeatability, the majority of samples used to establish and test the protocol were spiked with known viruses. However, in natural infections viruses are often cell-associated and large numbers of virus particles are intracellular. Therefore, sequencing of clinical samples from natural infections that showed positive results in specific qPCRs for different viruses were also necessary to test the virome protocol. Material from 7 samples (Tab.1.) was prepared according to the virome protocol (see 3.8) and sequenced on Illumina NextSeq 500 in FGCZ. The sequencing results (Tab. 3.) confirmed the specific qPCR results in all 7 cases. All viruses were successfully recovered by NGS. Additionally, the genome coverages of PCV-2 (3), SIV (5) and HEV (6 and 7) were 100%, providing the full-length genome sequences. The sequencing results of the undiagnosed pig brain sample with suspicion of porcine Teschovirus-1 infection showed no reads matching to the reference sequence of porcine Teschovirus-1, but 1000 reads matched to torque teno sus virus (TTSuV).

Number	Species	Material	Diagnosis by specific PCR	Number of reads matched to reference sequence database (NCBI)
1	bison	spleen	OvHV-2	10 501
2	sheep	skin crust	ORF virus	1540
3	pig	liver	PCV-2	21 000 000
4	pig	feces	PRV-A	218
5	pig	nasal swab	SIV N1H1	200 000
6	human	feces	HEV	119 000
7	Pig/Pork	sausage (mortadella di fegato	HEV	6 366

Tab. 3. Clinical samples tested by specific qPCR's in the Diagnostic department at the Institute of Virology in Vetsuisse-Faculty in Zürich and used to test the virome protocol.

4.10 Sensitivity of the protocol

To check the sensitivity of our NGS protocol in comparison to specific RT-qPCR, a SIV negative nasal swab sample was spiked with serial 10-fold dilutions of Influenza virus. The 1:10, 1:100 and 1:1000 dilutions resulted in 4806, 617 and 125 reads matching to the SIV reference genome (Fig.22). There was no Influenza virus detected in the 1:10'000 and 1:100'000 dilution, which by specific RT-qPCR were still weak positive with CT-values of 36.5 and 38.1, respectively.

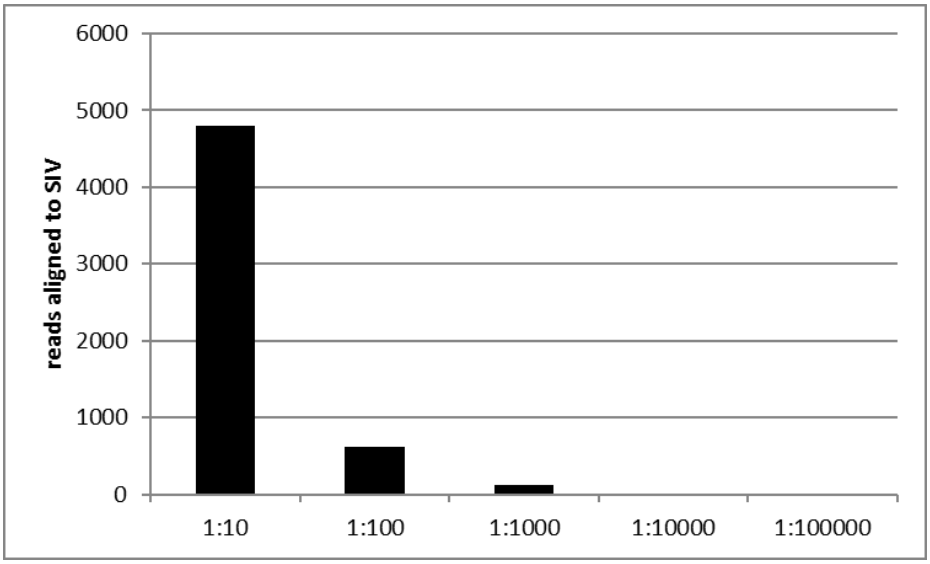

Fig. 22. Number of reads aligned to SIV reference sequence from a serial dilution of the sample spiked with SIV. In the 1:10'000 and 1:100'000 dilutions no Influenza was detected by NGS.

4.11 Full-length sequencing of Hepatitis E virus (HEV)

Using the virome protocol we successfully sequenced the full-length genomes of HEV from the stool sample of a human patient and the associated pork sausage. The detailed results can be found in the published genome announcement provided in Annex 1.

4.12 Whole genome sequencing and Swiss SIV surveillance program

The results of the comparison of targeted amplicon sequencing and the virome protocol (using different amounts of input material) for SIV whole genome sequencing is shown in Fig.23 and can be summarized as follows: all 8 segments of SIV were sequenced with all 4 methods. Targeted amplicon sequencing provided more reads from shorter segments like segment 7 or 8. The virome protocol provided more reads for longer segments such as 4 and 5. Furthermore, it worked most efficiently with the standard input volume, i.e., without concentration or using the double amount of sample material for nucleic acid isolation. All the segments of SIV were fully sequenced by all 4 methods, however the total number of reads aligned to reference was different (Fig. 23). These results show, that the virome protocol is applicable for SIV full genome sequencing directly form biological material and may replace specific RT-PCR for influenza typing.

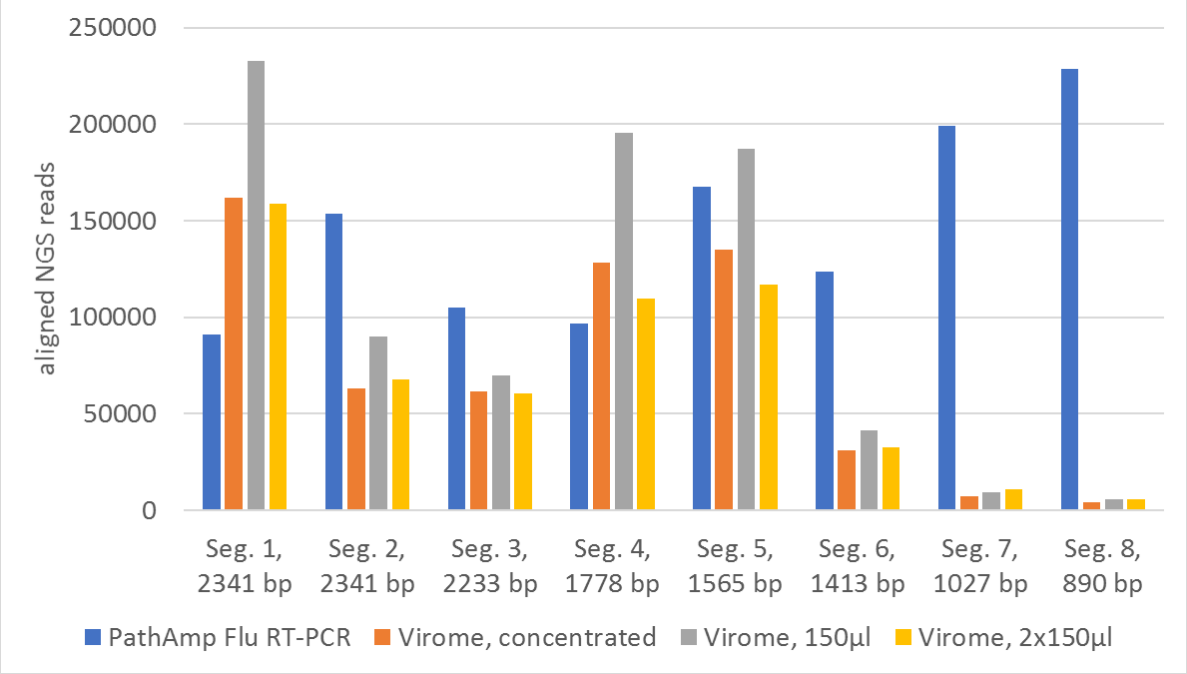

Fig. 23 Number of NGS reads aligned to SIV segments prepared by: PathAmp Flu kit (targeted amplicon sequencing), by virome protocol using concentrated samples, standard protocol and double amount of the sample input. The length of each segment is provided in base pairs (bp).

5. DISCUSSION

In this study, we made a first step to develop and establish a diagnostically applicable protocol for high throughput metagenomic sequencing of viruses in the veterinary field. Using NGS, all viruses that are present in the sample are sequenced, therefore there is no need to perform several specific tests. However, NGS is more time consuming than standard diagnostic tools and requires careful preparation of the samples. As reported by Daly et al, identification of viruses in human solid samples is technically challenging (Daly et al., 2011). It appears to be even more demanding for veterinary medicine laboratories since sample materials originate not only from various tissues, secretions and excretions but also from different species. Hence a flexible and versatile sample preparation method (e.g. working for solid and liquid samples) is required. Furthermore, a NGS protocol tailored to veterinary medicine has to be cost efficient particularly when applied in the livestock sector as financial resources are notoriously scarce and there may be high numbers of animals from the same farm that have to be tested.

So far, many reports of NGS analysis are based on porcine samples. Due to increased demand for pork meat, which is the most consummated worldwide, and with 40.1%, high growth rate and fertility, the swine industry is one of the most intensive animal production system all around the world. In addition, it has been forecasted that the need for pork will rise by 31% between 1997 and 2020 (Rosegrant et all., 2001). Taking those facts into account, infectious diseases affecting the pig population may have heavy impact and may lead to large economic losses for the swine industry. Moreover, pigs can act as a reservoir of zoonotic diseases, thereby posing a risk also for humans. Therefore, we chose to use mainly porcine sample material to establish our virome protocol. However, also samples from other species were used (Tab. 1) and successfully sequenced. Nevertheless, particularly the different consistency and composition of fecal samples from different species may require further testing.

The virome protocol presented here (Fig. 24) contains several stages of sample preparation that make it applicable for metagenomic studies. The first step is sample homogenization, an essential step in order to obtain a homogeneous solution for further sample processing. In solid samples, like tissue or leukocyte pellets, homogenization is particularly important in order to release the viruses from cells and make them accessible for further analysis. Since in metagenomic sequencing the whole DNA contained in a sample is subjected to sequencing, it is necessary to maximally reduce the host, bacterial and cellular DNA and RNA, without losing the viral nucleic acids. On the other hand, in fluid samples like saliva, blood or cell culture

supernatant, it may not have such a big influence (Daly et al., 2011). However, apart from viruses, homogenization also releases RNA/DNA of the host and bacteria and may damage viral particles. Therefore, it is necessary to find a balance between releasing viruses without damaging them but also reducing host nucleic acids. According to Conceição-Neto and colleagues, homogenization with ceramic beads reduces the number of viral particles, particularly coronavirus and mimivirus. Moreover, using beads may increase bacterial rRNA content in the sample (Conceição-Neto et al., 2015). According to our results, breaking down the tissue into very small pieces using a sterile scalpel blade, followed by intensive shaking in a buffered saline solution using the TissueLyser and subsequent centrifugation provided a similar amount of viral particles as using steel beads but without raising the total amount of host-derived DNA and RNA as much (see 4.1). However, the homogenization was tested using lung tissue and feces only. If skin samples or other more solid tissues are used, the steel beads may be required to allow proper homogenization.

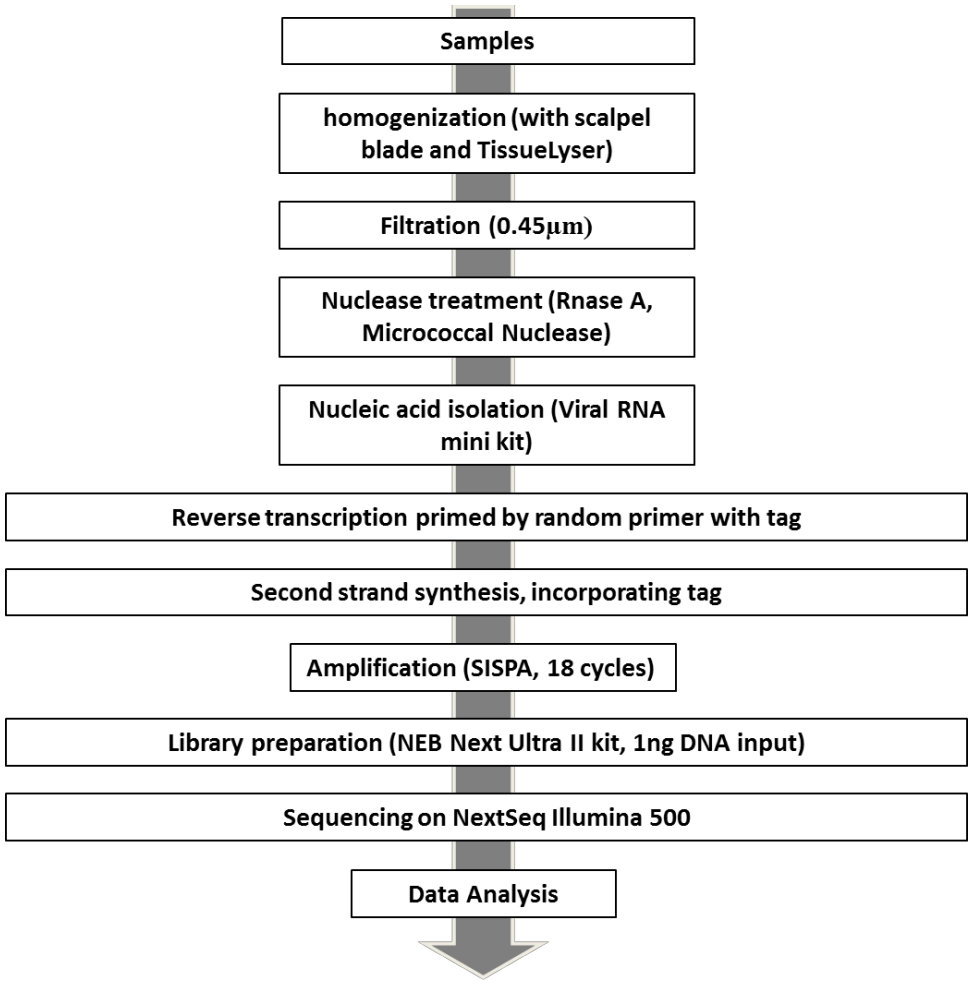

Fig. 24. Schematic representation of the optimized virome protocol.

Filtration is the second crucial step, that enables further separation of viral genomes from host and bacterial materials. As described before (Conceição-Neto et al., 2015), not only the filter pore size is important, but also filter pore material may influence filtration outcome. With two repetitions, we observed that there was no relevant influence on the CT values neither for BVDV nor BoHV-1 (Fig. 5). However, total DNA and RNA concentration was halved in samples filtrated using 0.45µm syringe filters compared to 0.8µm syringe filers (Fig. 6). Further analysis on more material and viruses may be necessary to draw a final conclusion. According to Kohl et al, 0.22µm filters provide the greatest host DNA reduction and the highest number of virus copies (Kohl et al., 2015). However, looking at the size range of bacteria and viruses, larger viruses such as pox virus and even herpesviruses may be reduced by 0.22 µm filtration. However, mycoplasma and other very small bacteria may well pass the 0.45 µm filters. Our successful application of the 0.45µm filters for parapoxvirus (Tab. 1) proves the method to work for large animal viruses.

Nuclease treatment is another important key point for the separation of viral nucleic acids from contaminating DNA and RNA. The idea is to degrade free nucleic acids while viral genomes are protected from the nucleases by their capsid. This depends also on how efficient host nucleic acids were released from the cells by homogenization and how intact the viral capsids are. The level of BVDV viral RNA and BoHV-1 DNA was nearly identical after treatment with all three nucleases and combinations thereof and was only slightly lower than without nuclease treatment (Fig. 7, 8 and 10). However, the combination of RNase A and Micrococcal Nuclease as well as RNase A and Benzonase or the combination of all three nucleases reduced the total RNA and DNA (and hence the contaminating nucleic acids) most efficiently (Fig. 9). Since the addition of Benzonase did not visibly improve the outcome and due to the relatively higher costs for Benzonase, we decided to include only treatment with RNase A and Micrococcal Nuclease. Some studies showed also that the 3-step combination of centrifugation, filtration (0,45µm syringe filter) and nuclease treatment (Hall et al., 2014), or the only 2-step combination with filtration and DNase treatment (Allander et al., 2001) provide significant relative increase in copies of viral genomes compared to the untreated sample. Importantly, the reduction of non-viral nucleic acids allows also reduction of sequencing depth – and therefore also sequencing costs.

Interestingly, both methods that were used to isolate nucleic acid are able to copurify RNA and DNA even if marketed for RNA extraction only. However, there was clearly less total DNA isolated with the viral RNA mini kit than with the peqGold TrifastFL (Fig. 12). In contrast, the viral RNA mini kit provided more viral DNA (lower BoHV-1 CT-values) than peqGold TrifastFL (Fig. 11), which was better for viral RNA isolation (lower BVDV CT-values). Overall, the viral RNA mini kit was chosen as the better test for our purpose, mainly because of its better co-purification of viral DNA and lower amount of host DNA. However, results from other studies, where more different virus familes were tested showed that TRIzol was more efficient for double-stranded RNA and DNA viruses, while Viral RNA mini kit was better for single-stranded RNA virueses (Kohl et al., 2015). Moreover, samples treated with TRIzol had less copies of host DNA, which in our results was opposite (Kohl et al., 2015). On the other hand Li et al. showed that extraction using viral RNA mini kit resulted in better nucleic acid isolation of all tested viruses than using TRIzol (Li et al., 2015). In contrast to other studies comparing extraction methods, in our case also the complete inactivation of the nucleases from the previous step by the first buffer of the extraction method was important. While this worked always well with TrifastFL, we had to add beta-mercaptoethanol to the AVL buffer of the viral

RNA mini kit to ensure complete nuclease inactivation (data not shown). In addition to its better performance, also the user friendliness and user safety plead in favour of the viral RNA mini kit, as it was easier and faster to use silka-columns. Moreover, health hazard arising from exposure to phenol and chloroform in the TRIzol protocol, can be avoided. Of course, many other kits and methods, such as magnetic bead systems, are available for nucleic acid extraction. However, due to limited finacial and time resources we focused on silica mambran and Phenol-chloroform based isolation only.

cDNA and 2nd strand synthesis is a required step for sequencing RNA and DNA viruses simultanelously. While, not so many researchers focus on this step we compared two kits: RevertAid First Strand H minus cDNA Synthesis kit and SuperScript IV Reverse Transcriptase. In all three spleen samples spiked with BVDV, more cDNA was obtained from samples prepared by RevertAid First Strand H minus cDNA Synthesis kit which is also much cheaper than SuperScript IV Reverse Transcriptase (Fig. 13).

Due to the massive reduction of total nucleic acids by viral particle enrichment, random amplification of the DNA is required in order to have enough DNA for further processing (library preparation). There are many different amplification methods possible such as the linker amplified shotgun library (LASL) or multiple displacement amplification (MDA) (Kim et al., 2011). Based on previous experience and as described before (Stalder et al., 2015), we have decided to use the SISPA method for random amplification. This method has been used (with variations) in many virome protocols (Li et al., 2015, Karlsson et al., 2013) and relies on integrating a known tag sequence during first and second strand synthesis followed by amplification with a single primer targeting the tag. While some researchers use separate workflows for RNA and DNA (Lewandowska et al., 2017) we chose a combined approach. Separate workflows mean that DNA is not affected by unnecessary reverse transcription. However, separate workflows will rise costs and time of sample preparation and we did not see an improvement in cDNA yield (data not shown). Since the kit used for library preparation is capable to process very low amounts of DNA (>= 0.5ng) and amplification is known to lead to relative overamplification of the most abundant DNA (which is usually host DNA), we tested to what extent we can reduce the number of amplification cycles in order to still have enough DNA for downstream application. While most authors use more than 30 cycles (Victoria et al., 2008, Ge et al., 2012, Stang et al., 2005, Kim and Bae 2011), we found 18 amplification cycles could efficiently amplify viral DNA that was successfully used for NGS (Fig. 14, 15 and 16). According to other studies, even 17 amplification cycles can be sufficient (Conceição-Neto et

al., 2015). Additionally, with less amplification cycles the non-viral DNA in the sample is less over-amplified and there is less danger of bias through polymerase errors.

Using two different library preparation kits with different amounts of input DNA, we observed that 1ng DNA input prepared by NEBnext Ultra II kit resulted in a higher percentage of reads assembled to the spiked-in viruses than the samples prepared from 50ng DNA input by NEBnext Ultra I and NEBNext Ultra II (Fig. 18). Due to these finding and as the DNA concentrations of our DNA samples ready for sequencing was generally relatively low, we decided to use the NEBnext Ultra II with only 1ng input for subsequent library preparations.

After establishing the different steps, we used the same starting material spiked with 6 different viruses and prepared i) without any enrichment, ii) only with filtration/nuclease treatment, iii) only with amplification or with both, iv) filtration/nuclease treatment AND amplification for a trial NGS run. The samples for which the full virome protocol was applied (filtration/nuclease treatment AND amplification) showed lower numbers of sequenced reads but a higher percentage of those reads were mapped to the spiked-in viruses (Fig. 19 and 20). This shows that in the wake of virus particle enrichment (homogenization, filtration and nucleases treatment), whole DNA background was successfully reduced while the spiked-in viruses remained untouched. Moreover, by lower numbers of reads in the full virome protocol sample, less false matched or non-relevant assemblings were generated and analysis time was reduced. In addition, all 6 spiked-in viruses were successfully recovered. Therefore, we concluded that the combination of enrichment for virus particles and subsequent random amplification is most promising regarding virus recovery and sequencing efficiency – and therefore also sequencing costs. This is also supported by the fact that similar protocols are frequently used for virome analyses by NGS (Victoria et al., 2009, Donaldson et al., 2010, Kohl et al., 2015).

The main goal of this study was to set a basis for the development of a diagnostically applicable NGS protocol virus diagnostics in veterinary medicine. Since virus particles spiked into e.g. a lung tissue may not entirely reflect a natural virus infection of the lung, where viruses are often intracellularly, clinical samples previously investigated by specific qPCR were used to examine if the protocol would also be applicable to “real” samples. To challenge the protocol, we used samples positive for a wide range of different RNA and DNA viruses, from some of the smallest viruses (porcine circovirus, ~20 nm) to some of the largest, pathogenic animal viruses (ovine parapoxvirus, 200-300nm), with double and single stranded genomes, with linear, segmented and circular genomes. Indeed, the viral genomes of all viruses detected by specific qPCR were also successfully detected by NGS (Tab. 3). Nevertheless, our results confirmed what other

virome studies have shown (Barzon et al., 2013): targeted diagnostic tests tend to have higher sensitivity than metagenomic approaches as we could not detect SIV diluted in nasal swab liquid at 1:10'000 and 100'000, corresponding to CT-values of 36.5 and 38.1, respectively (Fig. 22). In conclusion, for viral detection standard diagnostic methods and NGS should complement mutually, to have advantages of both methods i.e., rapidity and sensitivity of qPCR and high-throughput and primer independence of NGS when standard diagnostic fail in viral detection (Barzon et al., 2013).

As shown for SIV and HEV, our virome protocol can also be successfully used for sequencing of whole viral genomes (Bachofen et al., 2017). The consensus sequences of the HEV-3 were near full-length with only 39 and 69 nucleotides gaps in the ORF2 region that were subsequently bridged by Sanger sequencing. The consensus sequences of each SIV segments were full-length and revealed the same sequence as the PCR-based approach (Fig. 23). In contrast to PCR-based sequencing, also unknown viral genomes can be fully sequenced, e.g. of field strains. This is ideal for viruses with high evolutionary rates such as influenza virus, HIV or HCV, since all subtypes can be sequenced without need of using specific primers (Quiñones-Mateu et al., 2014). Hence, our NGS approach may be applied in national surveillance programs where sequence information is important, e.g. for the Swiss swine influenza virus surveillance program and could be helpful in molecular epidemiological aspects during outbreaks (Crisan et al., 2018).

In conclusion, the present study has demonstrated a way of samples preparation for various sample materials useable for metagenomics studies. Our 'virome protocol' sets a basis as a NGS approach for veterinary virology and has shown to be applicable to samples of multiple sources and for detection of a wide range of viruses. Based on the presented results, we recommend this protocol for the preparation of many animal samples used in routine virology diagnostics such as feces, tissue and swab samples. Furthermore, it could be used either for diagnostic purposes or research, e.g. for sequencing whole genomes and performing phylogenetic analyses.

We were only able to test a limited number of species and sample types during this thesis. Adaptations for specific species and sample types may be necessary and before metagenomics analysis can be applied routinely in veterinary diagnostic laboratories, further evaluation is needed. Standardization, relatively high sequencing costs and bioinformatic support as well as high computing resources and data storage, are main issues that need to be addressed. Nevertheless, we are confident that the virome protocol presented here will be implemented for

diagnostic purposes, particularly in cases of outbreaks of emerging, mutated or novel viral pathogens.

REFERENCES

Allander, T.; Emerson, S. U.; Engle, R. E.; Purcell, R. H.; Bukh, J. (2001): A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. In Proceedings of the National Academy of Sciences of the United States of America 98 (20), pp. 11609–11614.

Barzon, Luisa; Lavezzo, Enrico; Costanzi, Giulia; Franchin, Elisa; Toppo, Stefano; Palù, Giorgio (2013): Next-generation sequencing technologies in diagnostic virology. In Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology 58 (2), pp. 346–350.

Belák, S.; Karlsson, O. E.; Leijon, M.; Granberg, F. (2013): High-throughput sequencing in veterinary infection biology and diagnostics. In Revue scientifique et technique (International Office of Epizootics) 32 (3), pp. 893–915.

Bachofen C, Kubacki J, Jermini M, Giannini P, Martinetti G, Ripellino P, et al. New Subcluster of HEV Genotype 3 Strains Linked to the First Confirmed Swiss Case of Foodborne Hepatitis E Infection. Case Rep J. 2017;1(1):003

Blomström, Anne-Lie; Belák, Sándor; Fossum, Caroline; Fuxler, Lisbeth; Wallgren, Per; Berg, Mikael (2010): Studies of porcine circovirus type 2, porcine boca-like virus and torque teno virus indicate the presence of multiple viral infections in postweaning multisystemic wasting syndrome pigs. In Virus research 152 (1-2), pp. 59–64.

Boyle, David B.; Bulach, Dieter M.; Amos-Ritchie, Rachel; Adams, Mathew M.; Walker, Peter J.; Weir, Richard (2012): Genomic sequences of Australian bluetongue virus prototype serotypes reveal global relationships and possible routes of entry into Australia. In Journal of virology 86 (12), pp. 6724–6731.

Conceição-Neto, Nádia; Zeller, Mark; Lefrère, Hanne; Bruyn, Pieter de; Beller, Leen; Deboutte, Ward et al. (2015): Modular approach to customise sample preparation procedures for viral metagenomics. A reproducible protocol for virome analysis. In Scientific reports 5, p. 16532.

Crisan, Anamaria; McKee, Geoffrey; Munzner, Tamara; Gardy, Jennifer L. (2018): Evidence-based design and evaluation of a whole genome sequencing clinical report for the reference microbiology laboratory. In PeerJ 6, e4218.

Daly, Gordon M.; Bexfield, Nick; Heaney, Judith; Stubbs, Sam; Mayer, Antonia P.; Palser, Anne et al. (2011): A Viral Discovery Methodology for Clinical Biopsy Samples Utilising Massively Parallel Next Generation Sequencing. In PLoS ONE 6 (12).

Donaldson, Eric F.; Haskew, Aimee N.; Gates, J. Edward; Huynh, Jeremy; Moore, Clea J.; Frieman, Matthew B. (2010): Metagenomic analysis of the viromes of three North American bat species. Viral diversity among different bat species that share a common habitat. In Journal of virology 84 (24), pp. 13004–13018.

Eichwald, Catherine; Arnoldi, Francesca; Laimbacher, Andrea S.; Schraner, Elisabeth M.; Fraefel, Cornel; Wild, Peter et al. (2012): Rotavirus viroplasm fusion and perinuclear localization are dynamic processes requiring stabilized microtubules. In PLoS ONE 7 (10), e47947.

Feng, Huichen; Shuda, Masahiro; Chang, Yuan; Moore, Patrick S. (2008): Clonal integration of a polyomavirus in human Merkel cell carcinoma. In Science (New York, N.Y.) 319 (5866), pp. 1096–1100.

Ge, Xingyi; Li, Yan; Yang, Xinglou; Zhang, Huajun; Zhou, Peng; Zhang, Yunzhi; Shi, Zhengli (2012): Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. In Journal of virology 86 (8), pp. 4620–4630.

Hall, Richard J.; Wang, Jing; Todd, Angela K.; Bissielo, Ange B.; Yen, Seiha; Strydom, Hugo et al. (2014): Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. In Journal of virological methods 195, pp. 194–204.

Hatakeyama, Masaomi; Opitz, Lennart; Russo, Giancarlo; Qi, Weihong; Schlapbach, Ralph; Rehrauer, Hubert (2016): SUSHI. An exquisite recipe for fully documented, reproducible and reusable NGS data analysis. In BMC bioinformatics 17 (1), p. 228.

Hoffmann, Bernd; Scheuch, Matthias; Höper, Dirk; Jungblut, Ralf; Holsteg, Mark; Schirrmeier, Horst et al. (2012): Novel orthobunyavirus in Cattle, Europe, 2011. In Emerging infectious diseases 18 (3), pp. 469–472.

Johnson, Jeffrey A.; Geretti, Anna Maria (2010): Low-frequency HIV-1 drug resistance mutations can be clinically significant but must be interpreted with caution. In The Journal of antimicrobial chemotherapy 65 (7), pp. 1322–1326.

Karlsson, Oskar Erik; Belák, Sándor; Granberg, Fredrik (2013): The effect of preprocessing by sequence-independent, single-primer amplification (SISPA) on metagenomic detection of viruses. In Biosecurity and bioterrorism : biodefense strategy, practice, and science 11 Suppl 1, S227-34.

Kim, Kyoung-Ho; Bae, Jin-Woo (2011): Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. In Applied and environmental microbiology 77 (21), pp. 7663–7668.

Kohl, Claudia; Brinkmann, Annika; Dabrowski, Piotr W.; Radonić, Aleksandar; Nitsche, Andreas; Kurth, Andreas (2015): Protocol for metagenomic virus detection in clinical specimens. In Emerging infectious diseases 21 (1), pp. 48–57.

Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J. et al. (2001): Initial sequencing and analysis of the human genome. In Nature 409 (6822), pp. 860–921.

Lange, Jeannette; Groth, Marco; Schlegel, Michael; Krumbholz, Andi; Wieczorek, Kerstin; Ulrich, Roswitha et al. (2013): Reassortants of the pandemic (H1N1) 2009 virus and establishment of a novel porcine H1N2 influenza virus, lineage in Germany. In Veterinary microbiology 167 (3-4), pp. 345–356.

Lewandowska, Dagmara W.; Zagordi, Osvaldo; Geissberger, Fabienne-Desirée; Kufner, Verena; Schmutz, Stefan; Böni, Jürg et al. (2017): Optimization and validation of sample preparation for metagenomic sequencing of viruses in clinical samples. In Microbiome 5 (1), p. 94.

Li, Linlin; Deng, Xutao; Mee, Edward T.; Collot-Teixeira, Sophie; Anderson, Rob; Schepelmann, Silke et al. (2015): Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. In Journal of virological methods 213, pp. 139–146.

Masembe, Charles; Michuki, George; Onyango, Maria; Rumberia, Cecilia; Norling, Martin; Bishop, Richard P. et al. (2012): Viral metagenomics demonstrates that domestic pigs are a potential reservoir for Ndumu virus. In Virology journal 9, p. 218.

Quiñones-Mateu, Miguel E.; Avila, Santiago; Reyes-Teran, Gustavo; Martinez, Miguel A. (2014): Deep sequencing. Becoming a critical tool in clinical virology. In Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology 61 (1), pp. 9–19.

Rosegrant, Mark W.; Paisner, Michael S.; Meijer, Siet; Witcover, Julie (2001): 2020 global food outlook. Trends, alternatives, and choices. Washington, DC (A 2020 vision for food, agriculture, and the environment initiative).

Rushlow, K.; McNab, A.; Olson, K.; Maxwell, F.; Maxwell, I.; Stiegler, G. (1988): Nucleotide sequence of porcine rotavirus (OSU strain) gene segments 7, 8, and 9. In Nucleic acids research 16 (1), pp. 367–368.

Sanger, F.; Nicklen, S.; Coulson, A. R. (1977): DNA sequencing with chain-terminating inhibitors. In Proceedings of the National Academy of Sciences of the United States of America 74 (12), pp. 5463–5467.

Smith, Tara C.; Harper, Abby L.; Nair, Rajeshwari; Wardyn, Shylo E.; Hanson, Blake M.; Ferguson, Dwight D.; Dressler, Anne E. (2011): Emerging swine zoonoses. In Vector borne and zoonotic diseases (Larchmont, N.Y.) 11 (9), pp. 1225–1234.

Stalder, Hanspeter; Schweizer, Matthias; Bachofen, Claudia (2015): Complete Genome Sequence of a Bovine Viral Diarrhea Virus Subgenotype 1e Strain Isolated in Switzerland. In Genome announcements 3 (3).

Stang, Alexander; Korn, Klaus; Wildner, Oliver; Uberla, Klaus (2005): Characterization of virus isolates by particle-associated nucleic acid PCR. In Journal of clinical microbiology 43 (2), pp. 716–720.

Tebit, Denis M.; Arts, Eric J. (2011): Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. In The Lancet Infectious Diseases 11 (1), pp. 45–56.

Thomas, Torsten; Gilbert, Jack; Meyer, Folker (2012): Metagenomics - a guide from sampling to data analysis. In Microbial informatics and experimentation 2 (1), p. 3.

van Borm, Steven; Rosseel, Toon; Vangeluwe, Didier; Vandenbussche, Frank; van den Berg, Thierry; Lambrecht, Bénédicte (2012): Phylogeographic analysis of avian influenza viruses isolated from Charadriiformes in Belgium confirms intercontinental reassortment in gulls. In Archives of virology 157 (8), pp. 1509–1522.

van Dijk, Erwin L.; Jaszczyszyn, Yan; Thermes, Claude (2014): Library preparation methods for next-generation sequencing. Tone down the bias. In Experimental cell research 322 (1), pp. 12–20.

Victoria, Joseph G.; Kapoor, Amit; Dupuis, Kent; Schnurr, David P.; Delwart, Eric L. (2008): Rapid identification of known and new RNA viruses from animal tissues. In PLoS pathogens 4 (9), e1000163. DOI: 10.1371/journal.ppat.1000163.

Victoria, Joseph G.; Kapoor, Amit; Li, Linlin; Blinkova, Olga; Slikas, Beth; Wang, Chunlin et al. (2009): Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. In Journal of virology 83 (9), pp. 4642–4651. DOI: 10.1128/JVI.02301-08.

Wu, Zhiqiang; Ren, Xianwen; Yang, Li; Hu, Yongfeng; Yang, Jian; He, Guimei et al. (2012): Virome analysis for identification of novel mammalian viruses in bat species from Chinese provinces. In Journal of virology 86 (20), pp. 10999–11012.

Xu, Rui; Ekiert, Damian C.; Krause, Jens C.; Hai, Rong; Crowe, James E.; Wilson, Ian A. (2010): Structural basis of preexisting immunity to the 2009 H1N1 pandemic influenza virus. In Science (New York, N.Y.) 328 (5976), pp. 357–360.

Zhang, Xinsheng; Hasoksuz, Mustafa; Spiro, David; Halpin, Rebecca; Wang, Shiliang; Stollar, Sarah et al. (2007): Complete genomic sequences, a key residue in the spike protein and deletions in nonstructural protein 3b of US strains of the virulent and attenuated coronaviruses, transmissible gastroenteritis virus and porcine respiratory coronavirus. In Virology 358 (2), pp. 424–435.

Annex 1

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

VIRUSES

Complete Genome Sequences of Two Swiss Hepatitis E Virus Isolates from Human Stool and Raw Pork Sausage

Jakub Kubacki,[a] Cornel Fraefel,[a] Marco Jermini,[b] Petra Giannini,[b] Gladys Martinetti,[c] Paolo Ripellino,[d] Enos Bernasconi,[e] Xaver Sidler,[f] Roger Stephan,[g] Claudia Bachofen[a]

Institute of Virology, Vetsuisse Faculty University of Zurich, Zurich, Switzerland[a]; Department of Health, Cantonal Laboratory Ticino, Division of Public Health, Bellinzona, Switzerland[b]; Department of Laboratory Medicine, Laboratory of Bacteriology, Ente Ospedaliero Cantonale, Bellinzona, Switzerland[c]; Neurology Department, Neurocenter of Southern Switzerland, Ente Ospedaliero Cantonale, Bellinzona, Switzerland[d]; Internal Medicine and Infectious Diseases Department, Ente Ospedaliero Cantonale, Bellinzona, Switzerland[e]; Department of Farm Animals, Division of Swine Medicine, Vetsuisse Faculty University of Zurich, Zurich, Switzerland[f]; Institute for Food Safety and Hygiene, Vetsuisse Faculty University of Zurich, Zurich, Switzerland[g]

ABSTRACT We present here the full-length genome sequences of two hepatitis E virus genotype 3 (HEV-3) isolates from a human stool sample from a patient with acute hepatitis and a raw sausage containing pig liver. Sequence analysis implies that Swiss HEV isolates may form a novel subgroup of HEV-3 viruses.

Received 20 July 2017 **Accepted** 26 July 2017 **Published** 31 August 2017

Citation Kubacki J, Fraefel C, Jermini M, Giannini P, Martinetti G, Ripellino P, Bernasconi E, Sidler X, Stephan R, Bachofen C. 2017. Complete genome sequences of two Swiss hepatitis E virus isolates from human stool and raw pork sausage. Genome Announc 5:e00888-17. https://doi.org/10.1128/genomeA.00888-17.

Address correspondence to Claudia Bachofen, claudia.bachofen@uzh.ch.

Hepatitis E virus (HEV) is a major cause of acute hepatitis worldwide. It is a small nonenveloped or quasienveloped virus (1) with a positive-sense single-stranded RNA genome of approximately 7.2 kb and belongs to the family *Hepeviridae*. Genotypes 1 and 2 (HEV-1 and HEV-2, respectively) circulate within the human population and are a major health issue in developing countries. HEV-3 and HEV-4 are zoonotic viruses that are highly prevalent in porcine species and may be transmitted to humans by the consumption of pig liver and meat (2). HEV is also present in Switzerland (3, 4); however, aside from a single full-length sequence that was published recently (5), no further information on the diversity of HEV strains is available.

Here, we present the full-genome sequences of two HEV isolates from a human and a food sample. The stool sample originated in a 78-year-old male hospitalized in October 2016 in Lugano (Canton Ticino, southern Switzerland) with acute hepatitis and was collected within 10 days from symptom onset. Total RNA was extracted using the QIAamp viral RNA minikit, according to the manufacturer's instructions (Qiagen GmbH, Germany). The food sample was a traditional raw dry-cured pork sausage containing pork liver ("mortadella di fegato crudo"), sold by a local butcher shop. For the RNA extraction, a combination of TRI reagent (Lucerna-Chem AG, Luzern, Switzerland) and the NucliSENS easyMAG system (bioMérieux, Geneva, Switzerland) was used. Both RNAs were shown to be HEV positive by a commercial quantitative real-time reverse transcription-PCR (RT-PCR) (Ceeramtools [bioMérieux, Geneva, Switzerland]). To prepare the RNA samples for next-generation sequencing (NGS), sequence-independent single-primer amplification was performed (6), and the purified amplicons were used for the construction of sequencing libraries using the NEBNext Ultra II library preparation kit (BioConcept, Allschwil, Switzerland). A paired-end NGS run of 2 × 150 nucleotide read length was performed at the Functional Genomic Center Zurich using the Illumina NextSeq 500 machine. Alignment of the reads to full-length hepatitis E virus genomes using the SeqMan NGen software (DNAStar [Lasergene, Madison, WI, USA]) revealed best match to the recently published Swiss HEV-3 strain SW/16-0282 (GenBank accession no. KY780957 [5]). The consensus sequences of the alignments were near full

length, with only 39 and 69 nucleotide gaps in the open reading frame 2 (ORF2) region for the fecal sample and the sausage sample, respectively. The gaps were bridged by Sanger sequencing using specific primers binding upstream and downstream of the gaps. The 7,222-nucleotide full-length sequences [excluding the poly(A) tail] contain the 3 known HEV ORFs that are specified in detail in the GenBank entries. The sequences are identical except for 21 positions with nucleotide ambiguities that were confirmed by Sanger sequencing and point to quasispecies diversity. It is very likely that the two isolates belong to the same virus strain. Interestingly, this strain shows 95% identity to the only other fully sequenced Swiss isolate (accession no. KY780957 [5]) but only 88% identity to other HEV-3 strains. Analysis of more Swiss HEVs is necessary to confirm the existence of a Swiss-specific HEV-3 subcluster.

Accession number(s). Both sequences are deposited in GenBank under the accession numbers MF346772 and MF346773.

ACKNOWLEDGMENTS

This work was supported by internal funds from the Institute of Virology, Vetsuisse Faculty University of Zurich.

We thank the Functional Genomic Center Zurich for help with sequencing and analysis, Claudio Gobbi and Claudio Staedler from the Neurocenter of Southern Switzerland, Ospedale Civico, Lugano, for their clinical advice, and Lorenzo Leggeri from the Cantonal Laboratory, Bellinzona, for analytical support.

REFERENCES

1. Qi Y, Zhang F, Zhang L, Harrison TJ, Huang W, Zhao C, Kong W, Jiang C, Wang Y. 2015. Hepatitis E virus produced from cell culture has a lipid envelope. PLoS One 10:e1032503. https://doi.org/10.1371/journal.pone.0132503.
2. Salines M, Andraud M, Rose N. 2017. From the epidemiology of hepatitis E virus (HEV) within the swine reservoir to public health risk mitigation strategies: a comprehensive review. Vet Res 48:31. https://doi.org/10.1186/s13567-017-0436-3.
3. Burri C, Vial F, Ryser-Degiorgis MP, Schwermer H, Darling K, Reist M, Wu N, Beerli O, Schöning J, Cavassini M, Waldvogel A. 2014. Seroprevalence of hepatitis E virus in domestic pigs and wild boars in Switzerland. Zoonoses Publ Health 61:537–544. https://doi.org/10.1111/zph.12103.
4. Müller A, Collineau L, Stephan R, Müller A, Stärk KDC. 2017. Assessment of the risk of foodborne transmission and burden of hepatitis E in Switzerland. Int J Food Microbiol 242:107–115. https://doi.org/10.1016/j.ijfoodmicro.2016.11.018.
5. Wang B, Harms D, Hofmann J, Ciardo D, Kneubühl A, Bock C-T. 2017. Identification of a novel hepatitis E virus genotype 3 strain isolated from a chronic hepatitis E virus infection in a kidney transplant recipient in Switzerland. Genome Announc 5(20):e00345-17. https://doi.org/10.1128/genomeA.00345-17.
6. Daly GM, Bexfield N, Heaney J, Stubbs S, Mayer AP, Palser A, Kellam P, Drou N, Caccamo M, Tiley L, Alexander GJ, Bernal W, Heeney JL. 2011. A viral discovery methodology for clinical biopsy samples utilising massively parallel next generation sequencing. PLoS One 6:e28879. https://doi.org/10.1371/journal.pone.0028879.

Acknowledgements

This work was funded by the Federal Food Safety and Veterinary Office (BLV, project number 1.15.11).

I would like to thank everyone who contributed to the present dissertation, especially:

Dr. med. vet. Claudia Bachofen, PD, my supervisor in the Institute of Virology, Vetsuisse-Faculty, Zürich, for her great supervision and support.

Prof. Dr. Volker Thiel, Institute of Virology and Immunology, Vetsuisse-Faculty, Bern, for co-examining my dissertation.

Prof. Dr. Cornel Fraefel, the director of the Institute of Virology, Vetsuisse-Faculty, Zürich, for his scientific support.

Prof. Dr. Mathias Ackermann, former director of the Institute of Virology, Vetsuisse-Faculty, Zürich, for giving me the opportunity to work at the Institute and for his scientific support.

Members of our research group: Claudia Bachofen, Julia Lechmann, Karin Meier, Charlotte Rickli, Isabelle Specker, Marco Steck, Martina Buchser for all their help and feedback.

All the members of the Institute of Virology, Vetsuisse-Faculty, especially: Julia, Flo, Selim, Flo, Anna, Sameera, Neeta, Karin, Charlotte, Isabelle, Christina, Eva, with whom I had a great time while pursuing my doctoral project.

Finally, I would like to thank my family for all their support, especially Marta Nowak, for her feedback and patience.

Curriculum Vitae

Vorname Name	Jakub Kubacki
Geburtsdatum	19.03.1990
Geburtsort	Piła, Polen
Nationalität	Pole

September/1997 – Juni/2003	Primarschule, Piła, Polen
September/2003 – Juni/2006	Gymnasium, Piła, Polen
September/2006 – Juni/2009	Mittelschule, Piła, Polen

Oktober/2009 – März/2015	Studium der Veterinärmedizin, Universität von Ermland und Masuren, Olsztyn, Polen

24. März 2015	Abschlussprüfung vet. med., Universität von Ermland und Masuren, Olsztyn, Polen

Januar/2016 – April/2018	Anfertigung der Dissertation "A next generation sequencing approach for the diagnosis of outbreaks of emerging, mutated or novel viral pathogens" unter Leitung von Dr. med. vet. Claudia Bachofen, PD am virologischen Institut der Vetsuisse-Fakultät Universität Zürich Direktor Prof. Dr. Cornel Fraefel