Comparing Viral Metagenomic Extraction Methods

Jeanette Klenner[1,2], Claudia Kohl[1], Piotr W. Dabrowski[1,3] and Andreas Nitsche[1]*

[1]Centre for Biological Threats and Special Pathogens 1, Robert Koch Institute, Berlin, Germany.
[2]Bundeswehr Research Institute for Protective Technologies and NBC Protection, Munster, Germany.
[3]Research Group Bioinformatics (NG4), Robert Koch Institute, Berlin, Germany.

*Correspondence: nitschea@rki.de

https://doi.org/10.21775/cimb.024.059

Abstract

A crucial step in the molecular detection of viruses in clinical specimens is the efficient extraction of viral nucleic acids. The total yield of viral nucleic acid from a clinical specimen is dependent on the specimen's volume, the initial virus concentration and the effectiveness provided by the extraction method. Recent next generation sequencing (NGS)-based diagnostic approaches (i.e. metagenomics) provide a molecular 'open view' into the sample, as they theoretically generate sequence reads of any nucleic acid present in a specimen in a statistically representative manner. However, since a higher virus-related read output promises better sensitivity in the subsequent bioinformatic analysis, the extraction method selected determines the reliability of diagnostic NGS. In this study nine commercially available kits for nucleic acid extraction were compared regarding the simultaneous isolation of DNA and RNA by real-time PCR, four of which were selected for subsequent comparison by NGS (QIAamp Viral RNA Mini Kit, QIAamp DNA Blood Mini Kit, QIAamp cador Pathogen Mini Kit and QIAamp MinElute Virus Spin Kit). The nucleic acid yields and the sequence read output were compared for four different model viruses – reovirus, orthomyxovirus, orthopoxvirus and paramyxovirus – each at defined but varying concentrations in the same sample. The total amount of nucleic acid was processed to sequence the RNA (as cDNA) and the DNA with quantification by Qubit and virus-specific quantitative real-time PCRs. NGS libraries were prepared for sequencing on the Illumina HiSeq 1500 system. Finally, the percentage of reads assignable to each virus was determined via mapping.

Evaluation of different commercial nucleic acid extraction kits with four different viruses indicates little variation in the read numbers obtained for transcribed RNA or DNA by NGS.

Since NGS is increasingly being used as a tool in diagnostics of infectious diseases, the individual steps of the complete process have to be validated carefully. Here we could show

that for virus identification in liquid clinical specimens, any nucleic acid extraction kit that is performing well for PCR diagnostics can be used for NGS diagnostics as well and that the selection of the kit has only a minor impact on the yield of viral reads.

Introduction

The polymerase chain reaction (PCR) is currently the gold standard in nucleic acid-based diagnostics (Pabinger *et al.*, 2014; Schmittgen *et al.*, 2008). The multitude of rapid, highly specific and sensitive diagnostic PCR assays currently available for this well-understood and affordable method make it invaluable in the clinical context (Perandin *et al.*, 2004; Templeton *et al.*, 2004). However, the high specificity comes at a price: since PCR primers target a specific region of an organism's genome, the targeted pathogen is usually detected exclusively (Klein, 2002; Yamamoto, 2002).

A new method of nucleic acid-based analysis, next generation sequencing (NGS), has seen a soaring rise in popularity with several sequencing platforms developed independently of each other within less than a decade (Ahmadian and Svahn, 2011; Liu *et al.*, 2012; Metzker, 2010). This massively parallel sequencing approach provides a molecular 'open view' into the specimen which can allow for non-targeted pathogen detection (Hazelton and Gelderblom, 2003), thus overcoming the limitations set by PCR.

The application of NGS technologies (i.e. metagenomics) in the field of infectious diseases ranges from fundamental to applied research (Didelot *et al.*, 2012; Lecuit and Eloit, 2014; Mardis, 2008; Quail *et al.*, 2012; Renkema *et al.*, 2014). NGS is also increasingly used for diagnostics of bacterial and viral pathogens in different clinical specimen matrices, i.e. faeces, tissue, plasma, blood, urine, cerebrospinal fluid and diagnostic cell culture (Batty *et al.*, 2013; Cheval *et al.*, 2011; Kohl *et al.*, 2015; Law *et al.*, 2013; Nakamura *et al.*, 2008; Nassirpour *et al.*, 2014). It is generally assumed that NGS with its extremely rapid technical advancement and continuously decreasing costs will proceed to become a favourite tool in clinical routine diagnostics (Desai and Jere, 2012; Hayden, 2014; Peng *et al.*, 2013; Voelkerding *et al.*, 2009). The process of generating sequencing libraries and the technique of sequencing itself are well established. However, the impact of nucleic acid extraction from clinical samples on how well viral pathogens can be detected using metagenomics has not yet been comprehensively examined. Nevertheless, this is a critical question when considering adopting metagenomics as a standard method for clinical diagnostics.

As is generally accepted for clinical diagnostics, NGS-based diagnostics requires purification methods that are quick and allow for immediate sample processing, simple by not requiring specialized equipment and in the best case can be run automatically. Due to the open view approach it is of particular importance to isolate DNA and RNA simultaneously. To meet biosafety demands, reliable inactivation plays a critical role when dealing with clinical specimens (Boom *et al.*, 1990). In some kits the lysis buffer can contain chaotropic salts, i.e. guanidine hydrochloride/thiocyanate, which denature macromolecules, by which a virus inactivation is assumed to be likely (Blow *et al.*, 2004; Boom *et al.*, 1990).

Limiting factors in the nucleic acid extraction for NGS are specimen volume and low pathogen concentration that could result in insufficient amounts of NGS starting material or insufficient fragment length of the nucleic acid extracted (Fahle and Fischer, 2000). Several nucleic acid extraction kits are commercially available, but differ inter alia in cost, application, additional reagents required and hands-on effort. For PCR-based detection the

evaluation of several extraction kits for the recovery of pathogens such as *Y. pestis*, hepatitis A virus, mumps virus and cytomegalovirus has revealed that no kit is ideal under all conditions (Fahle and Fischer, 2000; Krause *et al.*, 2006; Peng *et al.*, 2013; Ribao *et al.*, 2004). However, such an evaluation is missing for NGS-based pathogen detection, i.e. metagenomics.

In this study we have therefore compared different commercially available RNA, DNA and combined RNA/DNA extraction kits regarding their performance when recovering four different model viruses present at different concentrations in fluid samples: a reovirus, an orthomyxovirus, an orthopoxvirus and a paramyxovirus. These model viruses were chosen in order to cover a wide range of different viral properties, i.e. non-enveloped particles, enveloped particles and single- or double-stranded genomic DNA or RNA. Samples were subjected to the identical diagnostic NGS process, and finally the number of reads mapping to the reference genome and the percentage of pathogen-specific reads in the total read amount obtained by sequencing on the Illumina HiSeq 1500 system were used as the performance measure.

Extraction kits

Eight commercially available kits [PureLink Viral RNA/DNA Kit (Invitrogen), ZR Viral DNA/RNA Kit (Zymonas), TRIzol LS Reagent (Invitrogen), QIAamp Viral RNA Mini Kit, QIAamp DNA Blood Mini Kit, QIAamp cador Pathogen Mini Kit, QIAamp MinElute Virus Spin Kit and QIAamp Ultrasense Kit (all Qiagen)] were compared for the simultaneous isolation of DNA and RNA, even for kits that are primarily designed for DNA or RNA exclusively. The selection of the individual kits was based on their commercial availability, the ease of handling and the total preparation time required. Therefore, the largest part of kits consists of silica spin columns, which turned out to be most efficient and popular for nucleic acid purification during the past decade.

Major differences of all kits utilized are in the different chaotropic salts included (i.e. guanidine hydrochloride/thiocyanate), detergents and other additives in the lysis buffers. In some kits the addition of protease as a separate additive is required. Some kits recommend the use of carrier DNA or RNA; however, in this study carrier DNA and RNA were omitted intentionally. Since typically NGS sequences all sequences in the specimen, carrier RNA or DNA would also be sequenced, and the viral sequencing read depth could be impaired. Pretests using quantitative real-time PCR were performed for nucleic acid extracts obtained from all kits (data not shown), and the four kits that scored well regarding the nucleic acid yields were chosen for further evaluation by using NGS (QIAamp Viral RNA Mini Kit, QIAamp DNA Blood Mini Kit, QIAamp cador Pathogen Mini Kit and QIAamp MinElute Virus Spin Kit). All four kits fulfilled the general requirements set: speed, simplicity, safety and no need for additional equipment. In contrast, all kits can be run manually or automatically on the QIAcube.

Artificial virus specimen

An aliquoted mix of four different RNA and DNA viruses was used as a well-defined surrogate for a liquid clinical specimen. This mix contained a reovirus (family *Reoviridae*, subfamily *Spinareovirinae*, genus *Orthoreovirus*, species Orthoreovirus T3/342/08), an Orthomyxovirus (family *Orthomyxoviridae*, genus *Influenzavirus A*, species H1N1 PR8/1934), an

Orthopoxvirus (family *Poxviridae*, subfamily *Chordopoxvirinae*, genus *Orthopoxvirus*, species vaccinia virus) from cell culture supernatant and a paramyxovirus (family *Paramyxoviridae*, subfamily *Paramyxovirinae*, genus *Respirovirus*, species Sendai virus) from the allantoic fluid of an infected egg. To produce viruses from cell culture supernatant, Vero E6 cells together with the respective virus were grown to confluence in DMEM with 10% Fetal Bovine Serum (FCS) and 2 mM L-glutamine. Viral supernatant was collected in a tube on day five and centrifuged at 200 rpm for 10 minutes, and the pellet was resuspended in FCS. For Paramyxovirus propagation allantoic fluid of infected chicken eggs was handled as previously described by Kohl *et al.* (2015). Each virus was present at defined but varying concentrations in the same sample (as determined by quantitative real-time PCR). A whole aliquot of the mix was used for each extraction.

NGS workflow

The workflow used to compare the different extraction kits' performance on the selected viruses is shown in Fig. 4.1 and described in detail below. Compared to PCR, NGS is still

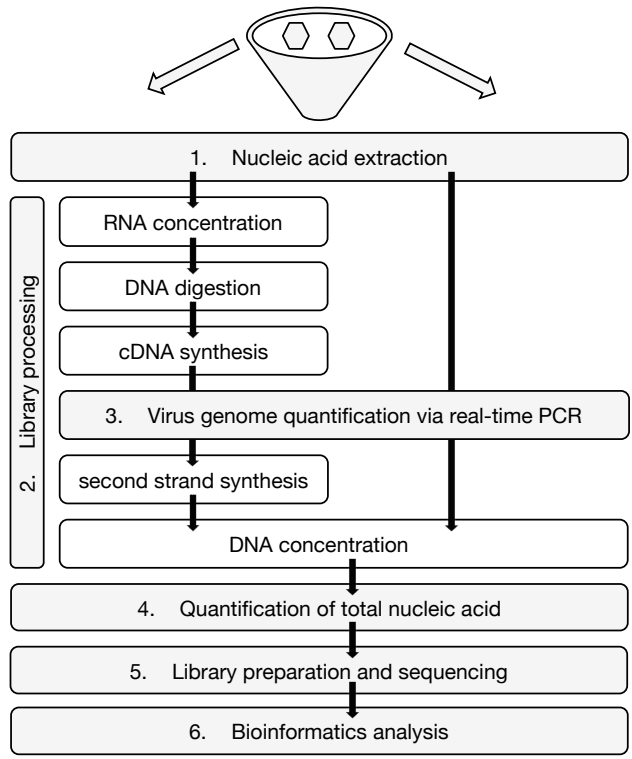

Figure 4.1 Experimental workflow. After total nucleic acid extraction, each sample was divided into two aliquots. Each of the aliquots was processed to either DNA or RNA. For the latter, DNA digestion, cDNA synthesis and second strand synthesis were performed in addition. The DNA and the RNA aliquot were further treated identically. Details of each step (1–6) of the workflow are given in the text under the corresponding section title.

an expensive method, despite the steadily decreasing costs. Therefore, the NGS runs were performed in duplicate, as a higher number of replicates would have substantially increased the cost of the study.

Nucleic acid extraction

Nucleic acid was extracted in duplicate following the manufacturers' instructions, except that no carrier RNA/DNA was used for the reasons mentioned above (Table 4.1). The final DNA and RNA solutions were eluted in the same volume of 100 µl of AE or AVE buffer (Qiagen), respectively.

Library processing

Each nucleic acid solution was divided into two aliquots, with one of the aliquots being further subjected to RNA and the other to DNA processing for NGS.

One aliquot was concentrated to a final volume of 13 µl using the RNeasy MinElute Cleanup Kit (Qiagen) and subjected to DNase digestion by Turbo DNA-free (Life Technologies, Darmstadt, Germany) at 37°C for 20 minutes according to the manufacturers' instructions. Total purified and concentrated RNA was reverse-transcribed into cDNA using random hexamers, RNaseOUT™ and Superscript II (Life Technologies) following the manufacturers' instructions, albeit an additional denaturing step of 95°C for 5 minutes

Table 4.1 Comparison of the different extraction kits (according to the manufacturer's information)

Kit[a]	Target	Specimen type	Costs per kit[b] (€)	Requirements[c]		Starting volume (µl)	Elution volume (µl)
				Reagents	Special equipment[d]		
VRMK	Viral RNA	Cell-free body fluids	197	Ethanol[e]	Heating block, Vortex, Microcentrifuge	140	60–80[f]
DBMK	Viral DNA	Blood and related body fluids	145	Ethanol[e], PBS[g]	Heating block, Vortex, Microcentrifuge	200	50–200[h]
CPMK	Viral RNA, viral DNA Bacterial DNA	Whole blood, serum and tissue	176	Ethanol[e], PBS[g]	Vortex, Microcentrifuge	200	50–150
MEVK	Viral RNA, Viral DNA	Plasma, serum and cell-free body fluids	224	Ethanol[e], PBS[g]	Heating block, Vortex, Microcentrifuge	200	20–150

[a]By PCR pre-selected kits. VRMK, QIAamp Viral RNA Mini Kit; DBMK, QIAamp DNA Blood Mini Kit; CPMK, QIAamp cador Pathogen Mini Kit; MEVK, QIAamp MinElute Virus Spin Kit.
[b]Pricing quote for small-volume orders (50), Germany, as of April 2016 (€).
[c]Not included in the kit.
[d]Additional laboratory equipment.
[e]Use 96–100% ethanol.
[f]A single elution with 60 µl of buffer AVE is sufficient to elute at least 90% of the viral RNA. Performing a double elution using 2 × 40 µl of buffer AVE will increase yield by up to 10%.
[g]Phosphate-buffered saline (PBS) may be required for some samples.
[h]If more eluate is required increase the amount of buffer AVE used in the two elution steps (2 × 50 µl instead of 2 × 30 µl).

was used for initial RNA denaturation. Second strand synthesis was performed with the NEBNext® Second Strand Synthesis Module according to the manufacturers' instructions (New England Biolabs, Frankfurt/Main, Germany). The resulting double-stranded cDNA (ds-cDNA) and the originally extracted DNA fraction were further purified with the MinElute PCR Purification Kit (Qiagen). In none of the experiments was the capacity limit of the columns of 5 µg exceeded (Table 4.2).

Virus genome quantification via real-time PCR

Following nucleic acid extraction by the different kits (Table 4.1), their individual performance regarding the yield of viral nucleic acids was compared by quantitative real-time PCR. RNA was reversely transcribed into cDNA as described in the paragraph above. Specific real-time PCR protocols were applied in duplicate to each aliquot, using the Applied Biosystems 7500 Real Time PCR System (Applied Biosystems, Darmstadt, Germany), for the quantification of T3/Bat/Germany/342/08 Reovirus as previously described by Kohl *et al.* (2012), Orthomyxovirus/Influenza viruses by Schulze *et al.* (2010), Orthopoxviruses by Schröder and Nitsche (2010) and Sendai virus by Kohl *et al.* (2015). The PCR reaction mixture consisted of 1 U Platinum Taq DNA polymerase (Invitrogen, Darmstadt, Germany), 1x Platinum Taq Buffer (Invitrogen), 5 mM $MgCl_2$ (Invitrogen), 300 nM of each primer, 100 nM of TaqManprobe and 100 µM mix of dNTPs (Invitrogen). All PCR reaction mixtures were performed in a final volume of 25 µl containing 3 µl of the samples. Identical conditions were used for all PCR reactions: an enzyme activation step at 95°C for 10 minutes, followed by 45 cycles with a 95°C denaturation step for 15 s and 60°C annealing for 35 s (Bustin *et al.*, 2009).

Quantification of total nucleic acid

Prior to further library processing, the yield of transcribed RNA and DNA extracted was determined by Qubit® 2.0 Fluorometer (Qubit® dsDNA HS Assay Kit, Invitrogen). The assay is designed for the accurate measurement of sample concentrations from 10 pg/µl to 100 ng/µl.

Table 4.2 Nucleic acid concentration of samples with different kits, determined by Qubit (mean concentration, n=2)

	Concentration (ng/µl)	
Kits[a]	RNA[b]	DNA[c]
VRMK	b/d	2.3
DBMK	b/d	1.0
CPMK	b/d	1.0
MEVK	b/d	2.2

[a]By PCR pre-selected kits. VRMK, QIAamp Viral RNA Mini Kit; DBMK, QIAamp DNA Blood Mini Kit; CPMK, QIAamp cador Pathogen Mini Kit; MEVK, QIAamp MinElute Virus Spin Kit.
[b]After RNA preparation (ds-cDNA).
[c]After DNA preparation.
b/d, below Qubit detection limit.

Library preparation and sequencing

NGS libraries were generated with the Nextera XT DNA Sample Preparation Kit according to the manufacturer's instructions (Illumina, San Diego, CA, USA). For quantification of NGS libraries the KAPA Library Quantification Kits for Illumina sequencing (Kapa Biosystems, Wilmington, MA, USA) were used. Due to the fact that the starting amount of 1 ng of nucleic acid was not reached in the case of the RNA preparations after the second strand synthesis, the entire sample volume was added to the library. Both the RNA and DNA libraries were sequenced. Sequencing was performed on the Illumina HiSeq 1500 system in rapid-run mode. A paired-end protocol was used with read lengths of 2 × 150 bp and a total run time of 40 h.

Bioinformatics analysis

The bioinformatics analysis was performed by mapping the reads obtained to the reference genome sequences of the four viruses present in the sample, using bowtie2 (Langmead and Salzberg, 2012). Reference sequences were: reovirus: NCBI accession numbers JQ412763.1, JQ412761.1, JQ412759.1, JQ412757.1, JQ412755.1, JQ412764.1, JQ412762.1, JQ412760.1, JQ412758.1 and JQ412756.1; influenza virus: NCBI accession numbers NC_002018.1, NC_002016.1, NC_002023.1, NC_002021.1, NC_002019.1, NC_002017.1, NC_002020.1 and CY033582.1; Orthopoxvirus: NCBI accession number NC_006998.1; and Sendai virus: NCBI accession number EF679198.1.

Comparison of extraction kit performance

For the comparison of different commercially available DNA and RNA extraction kits, four different viruses with different characteristics were chosen: Reovirus, Orthomyxovirus, Orthopoxvirus and Paramyxovirus. Following nucleic acid extraction, NGS was applied in order to determine the performance of each kit regarding viral nucleic acid detection.

The nucleic acid concentrations following DNA preparation were close to 1 ng/µl which is the amount required theoretically for library preparation with the Nextera XT Sample Prep Kit® according to the manufacturer's instructions (Table 4.2). The highest DNA concentration on average was found in samples extracted by the QIAamp Viral RNA Mini Kit (2.3 ng/µl), while extraction with the QIAamp DNA Blood Mini Kit and QIAamp cador pathogen detection kit resulted in the lowest concentration (1.0 ng/µl). However, all four viruses were detectable after nucleic acid preparation by quantitative real-time PCR, indicating a detectable amount of viral nucleic acid in the respective samples (Table 4.3). Interestingly, both DNA and RNA viruses were detected using the QIAamp Viral RNA Mini Kit and the QIAamp DNA Blood Mini Kit, although these kits are designed for either DNA or RNA.

According to quantitative real-time PCR results, the efficiency of Orthomyxovirus, Orthopoxvirus and Orthoreovirus genome extraction was higher when using the QIAamp cador Pathogen Mini Kit than when using the QIAamp MinElute Virus Spin Kit. The lowest C_T values for Paramyxovirus were reached for samples extracted with the QIAamp Viral RNA Mini Kit (C_T 15.8 ± 0.4) while the QIAamp MinElute Virus Spin Kit had the lowest detectability (C_T 17.4).

Illumina sequencing of the respective libraries following DNA or RNA preparation resulted in 93 million paired-end reads (150 + 150 bp), with a total of 24.5 Gbp sequence

Table 4.3 Comparison of different extraction kits based on average threshold cycle (C_T) during real-time PCR (n=2, CI=95%). The best result is presented in bold

	Orthomyxovirus	Orthopoxvirus	Orthoreovirus	Paramyxovirus
Kits[a]	RNA genome[b]	DNA genome[c]	RNA genome[b]	RNA genome[b]
VRMK	**20.9 ± 0.2**	17.6 ± 0.2	28.8 ± 1.1	**15.8 ± 0.4**
DBMK	21.5 ± 0.2	18.0 ± 0.2	29.8 ± 0.2	16.3 ± 0.5
CPMK	21.2 ± 0.4	**17.5 ± 0.2**	**28.7 ± 1.1**	16.3 ± 1.2
MEVK	21.6[d]	18.9 ± 0.2	30.4[d]	17.4[d]

[a]By PCR pre-selected kits. VRMK, QIAamp Viral RNA Mini Kit; DBMK, QIAamp DNA Blood Mini Kit; CPMK, QIAamp cador Pathogen Mini Kit; MEVK, QIAamp MinElute Virus Spin Kit.
[b]After RNA preparation (ds-cDNA).
[c]After DNA preparation.
[d]Single measurement due to insufficient sample volume.

information. On average, the percentage of bases with a quality score greater than 30 was 88.15%. The average number of reads from each library and the percentage of reads mapping to the reference genomes of all viruses examined are shown in Table 4.4.

In both libraries derived from DNA and RNA preparation, the number of non-viral reads exceeded the number of viral reads. The total number of sequence reads for each sample ranged from 5,827,117 to 10,537,950 reads in the RNA preparation and from 9,176,541 to 13,256,053 reads in the DNA preparation. On average, a higher proportion of non-viral reads was observed following DNA preparation (98.92%) compared to RNA preparation (71.8%).

Overall, the QIAamp MinElute Virus Spin Kit allowed the recovery of the highest percentage of viral reads after RNA preparation while generating the lowest percentage of non-viral reads. The QIAamp cador Pathogen Mini Kit showed the second-best performance regarding RNA preparation read recovery and the best read recovery for the Orthopoxvirus DNA library. The largest difference in percentage of recovered reads was observed in the Orthopoxvirus library, where the highest percentage was close to twofold higher than the lowest percentage. As visualized in Fig. 4.2, the performance of the kits concerning the tested viruses generally showed only small differences, and all viruses could be detected via NGS after preparation with each kit.

As shown in Table 4.3, the results obtained from real-time PCR are somewhat different. The QIAamp MinElute Virus Spin Kit, which allowed for the recovery of the highest percentage of viral reads for RNA viruses and an average percentage of viral reads for the DNA virus by NGS, produced the highest C_T values for all viruses. However, the QIAamp cador Pathogen Mini Kit produced both the highest percentage of Orthopoxvirus reads in NGS and the lowest Orthopoxvirus C_T of all kits tested.

Discussion

In this study we compared four pre-selected QIAGEN kits (QIAamp Viral RNA Mini Kit, QIAamp DNA Blood Mini Kit, QIAamp cador Pathogen Mini Kit and QIAamp MinElute Virus Spin Kit) with NGS in terms of their ability to recover sequence information of four different viruses (Reovirus, Orthomyxovirus, Orthopoxvirus and Paramyxovirus). These

Table 4.4 Percentage (%) of viral reads and total number of reads (#) obtained after Illumina sequencing. The best results are given in bold

	Orthomyxovirus	Orthopoxvirus	Orthoreovirus	Paramyxovirus	Non-viral reads		No. of reads[d]	
Kit[a]	RNA[b] (%)	DNA[c] (%)	RNA[b] (%)	RNA[b] (%)	RNA[b] (%)	DNA[c] (%)	RNA[b]	DNA[c]
VRMK	6.7	0.7	0.001	17.9	75.4	99.3	**1.05×10^7**	**1.33×10^7**
DBMK	5.8	1.0	0.001	16.6	77.6	99.0	6.11×10^6	1.21×10^7
CPMK	6.9	**1.6**	0.003	23.4	69.6	98.4	7.86×10^6	1.01×10^7
MEVK	**11.3**	1.0	0.003	**24.0**	64.6	99.0	5.83×10^6	9.18×10^6

[a]By PCR pre-selected kits. VRMK, QIAamp Viral RNA Mini Kit; DBMK, QIAamp DNA Blood Mini Kit; CPMK, QIAamp cador Pathogen Mini Kit; MEVK, QIAamp MinElute Virus Spin Kit.
[b]% of reads after RNA preparation (ds-cDNA).
[c]% of reads after DNA preparation.

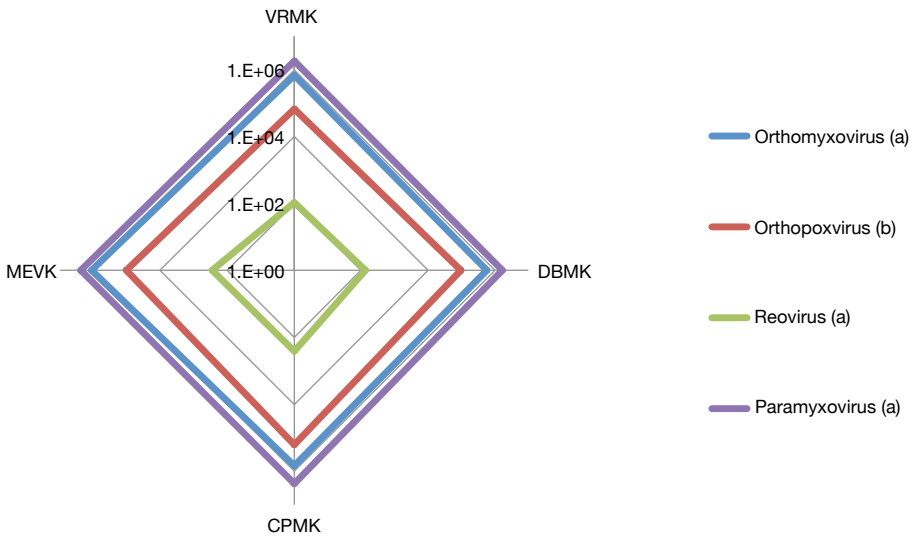

Figure 4.2 Number of viral reads recovered after extraction of samples using four different extraction kits. The viruses used were Reovirus (T3/Bat/Germany/342/08), Orthomyxovirus (H1N1 PR8/34), Orthopoxvirus (Vaccinia Virus) and Paramyxovirus (Sendai virus). Recovered viral reads were normalized to 10 million output reads. Normalized read numbers are shown in log scale. VRMK: QIAamp Viral RNA Mini Kit, DBMK: QIAamp DNA Blood Mini Kit, CPMK: QIAamp cador Pathogen Mini Kit and MEVK: QIAamp MinElute Virus Spin Kit. (a) after RNA preparation (ds-cDNA), (b) after DNA preparation.

viruses with varying genome characteristics were selected to cover a broad range of viruses possibly present in a clinical sample. In contrast to a clinical sample, the mixed aliquots were all at fairly high titres with a minor complexity because the focus of this study was the comparison of different extraction kits and not the evaluation of the detection limit of NGS for metagenomics. While similar studies have shown that for PCR – the current gold standard for nucleic acid-based diagnostics – no single kit is perfect for all pathogens (Dauphin *et al.*, 2010, Fahle and Fischer, 2000; Krause *et al.*, 2006; Ribao *et al.*, 2004), to our knowledge this is the first study directed at NGS-based pathogen detection.

As shown in Table 4.4, all viruses could be detected using any of the kits compared, with the highest percentage of reads recovered by the QIAamp MinElute Virus Spin Kit for Reovirus, Orthomyxovirus and Paramyxovirus and by the QIAamp cador Pathogen Mini Kit for the Orthopoxviruses, respectively. Interestingly, this is in contrast to the results (C_T values) obtained for each recovery by corresponding real-time PCR shown in Table 4.3. In the real-time PCR results, QIAamp MinElute Virus Spin Kit has the highest C_T value for all viruses tested. PCR quantifies the total number of target sequences in the sample, while NGS quantifies the number of target sequences in relation to the total number of sequences in the sample. Thus, when preparing nucleic acid for PCR-based detection, it is most important not to decrease the total amount of nucleic acid in the sample to preserve as many total PCR targets as possible; however, dilution of the nucleic acid prior to PCR reaction may sometimes be required to prevent PCR inhibition by too high amounts of DNA.

In contrast, when preparing nucleic acid for NGS-based detection (i.e. metagenomics), it is most important to deplete the host nucleic acid more efficiently than the virus nucleic acid. It is to a certain degree acceptable to decrease the pathogen nucleic acid, as long as the host nucleic acid is decreased by a higher factor to increase the detection likelihood. Subsequently, available publications comparing different preparation kits, based on their performance in PCR-based detection methods alone, are not expedient for NGS-based detection approaches.

In the study presented, only slight differences in the number of recovered NGS reads were observed between all kits tested – the highest difference observed between the recoveries of two kits is just above twofold (Fig. 4.2). While not conclusive due to the defined number of model organisms tested, our results suggest that NGS-based detection is not dependent on the choice of nucleic acid extraction kit.

This allows choosing kits according to other parameters in the decision process. One factor could be costs which range from €2.90 to €4.48 per reaction, the most economical solution being the QIAamp Viral RNA Mini Kit. Or the QIAamp cador Pathogen Mini Kit is the only one tested not requiring a heating block. All of the kits tested are compatible with the QIAcube automated extraction system. In summary, while all kits perform similarly well, the QIAamp MinElute Virus Spin Kit seems most advantageous for NGS due to the low elution volume of 20 µl. In general, detection of viruses is made difficult by their structural and genetic diversity (Hulo *et al.*, 2011). The lack of a universally conserved region analogous to bacterias' 16S rRNA prohibits generic amplification and enrichment. The small size of the viral genome compared to the host genome (e.g. the porcine circoviruses, with about 1.7 kb, versus the human genome, with 3,000,000 kb) adversely affects the nucleic acid ratio in a sample (Finsterbusch and Mankertz, 2009; Hulo *et al.*, 2011; Nagele, 2011). Background reads such as the human genome thus present a major problem when using NGS for pathogen detection, necessitating to rely either on huge sequencing depth or on luck for finding viral sequences. This attaches importance to the methods for extraction and enrichment of viral nucleic acid from clinical samples which allow better detection of pathogens using NGS; obviously an important factor in the adoption of NGS as a standard method for clinical diagnostics. Available research data show that in metagenomic and virome studies the virus purification and enrichment have a high level of importance. Optimization of the extraction process has already led to an increase in the detection likelihood of viruses from organ tissue (Kohl *et al.*, 2015). In summary, one can use any of the compared kits for the extraction of nucleic acid from fluid samples.

Acknowledgements

The authors are grateful to Dr Brunhilde Schweiger for providing Influenza A virus strain PR8/38, Dr Marc Hoferer and Dr Andreas Kurth for providing the Sendai virus isolate and Ursula Erikli for copy-editing.

References

Ahmadian, A., and Svahn, H.A. (2011). Massively parallel sequencing platforms using lab on a chip technologies. Lab Chip *11*, 2653–2655. http://dx.doi.org/10.1039/c1lc90035h

Batty, E.M., Wong, T.H., Trebes, A., Argoud, K., Attar, M., Buck, D., Ip, C.L., Golubchik, T., Cule, M., Bowden, R., *et al.* (2013). A modified RNA-Seq approach for whole genome sequencing of RNA viruses from faecal and blood samples. PLOS ONE *8*, e66129. http://dx.doi.org/10.1371/journal.pone.0066129

Blow, J.A., Dohm, D.J., Negley, D.L., and Mores, C.N. (2004). Virus inactivation by nucleic acid extraction reagents. J. Virol. Methods *119*, 195–198. http://dx.doi.org/10.1016/j.jviromet.2004.03.015

Boom, R., Sol, C.J., Salimans, M.M., Jansen, C.L., Wertheim-van Dillen, P.M., and van der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. J. Clin. Microbiol. *28*, 495–503.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., *et al.* (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. *55*, 611-622.

Cheval, J., Sauvage, V., Frangeul, L., Dacheux, L., Guigon, G., Dumey, N., Pariente, K., Rousseaux, C., Dorange, F., Berthet, N., *et al.* (2011). Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. J. Clin. Microbiol. *49*, 3268–3275. http://dx.doi.org/10.1128/JCM.00850-11

Dauphin, L.A., Stephens, K.W., Eufinger, S.C., and Bowen, M.D. (2010). Comparison of five commercial DNA extraction kits for the recovery of Yersinia pestis DNA from bacterial suspensions and spiked environmental samples. J. Appl. Microbiol. *108*, 163–172. http://dx.doi.org/10.1111/j.1365-2672.2009.04404.x

Desai, A.N., and Jere, A. (2012). Next-generation sequencing: ready for the clinics? Clin. Genet. *81*, 503–510. http://dx.doi.org/10.1111/j.1399-0004.2012.01865.x

Didelot, X., Bowden, R., Wilson, D.J., Peto, T.E., and Crook, D.W. (2012). Transforming clinical microbiology with bacterial genome sequencing. Nat. Rev. Genet. *13*, 601–612. http://dx.doi.org/10.1038/nrg3226

Fahle, G.A., and Fischer, S.H. (2000). Comparison of six commercial DNA extraction kits for recovery of cytomegalovirus DNA from spiked human specimens. J. Clin. Microbiol. *38*, 3860–3863.

Finsterbusch, T., and Mankertz, A. (2009). Porcine circoviruses – small but powerful. Virus Res. *143*, 177–183. http://dx.doi.org/10.1016/j.virusres.2009.02.009

Hayden, E.C. (2014). Technology: The $1,000 genome. Nature *507*, 294–295. http://dx.doi.org/10.1038/507294a

Hazelton, P.R., and Gelderblom, H.R. (2003). Electron microscopy for rapid diagnosis of infectious agents in emergent situations. Emerg. Infect. Dis. *9*, 294–303.

Hulo, C., de Castro, E., Masson, P., Bougueleret, L., Bairoch, A., Xenarios, I., and Le Mercier, P. (2011). ViralZone: a knowledge resource to understand virus diversity. Nucleic Acids Res. *39*, D576–82. http://dx.doi.org/10.1093/nar/gkq901

Klein, D. (2002). Quantification using real-time PCR technology: applications and limitations. Trends Mol. Med. *8*, 257–260.

Kohl, C., Lesnik, R., Brinkmann, A., Ebinger, A., Radonić, A., Nitsche, A., Mühldorfer, K., Wibbelt, G., and Kurth, A. (2012). Isolation and characterization of three mammalian orthoreoviruses from European bats. PLOS ONE *7*, e43106. http://dx.doi.org/10.1371/journal.pone.0043106

Kohl, C., Brinkmann, A., Dabrowski, P.W., Radonić, A., Nitsche, A., and Kurth, A. (2015). Protocol for metagenomic virus detection in clinical specimens. Emerg. Infect. Dis. *21*, 48–57. http://dx.doi.org/10.3201/eid2101.140766

Krause, C.H., Eastick, K., and Ogilvie, M.M. (2006). Real-time PCR for mumps diagnosis on clinical specimens – comparison with results of conventional methods of virus detection and nested PCR. J. Clin. Virol. *37*, 184–189.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. http://dx.doi.org/10.1038/nmeth.1923

Law, J., Jovel, J., Patterson, J., Ford, G., O'Keefe, S., Wang, W., Meng, B., Song, D., Zhang, Y., Tian, Z., *et al.* (2013). Identification of hepatotropic viruses from plasma using deep sequencing: a next generation diagnostic tool. PLOS ONE *8*, e60595. http://dx.doi.org/10.1371/journal.pone.0060595

Lecuit, M., and Eloit, M. (2014). The diagnosis of infectious diseases by whole genome next generation sequencing: a new era is opening. Front. Cell. Infect. Microbiol. *4*, 25. http://dx.doi.org/10.3389/fcimb.2014.00025

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. (2012). Comparison of next-generation sequencing systems. J. Biomed. Biotechnol. *2012*, 251364. http://dx.doi.org/10.1155/2012/251364

Mardis, E.R. (2008). The impact of next-generation sequencing technology on genetics. Trends Genet. *24*, 133–141. http://dx.doi.org/10.1016/j.tig.2007.12.007

Metzker, M.L. (2010). Sequencing technologies - the next generation. Nat. Rev. Genet. *11*, 31–46. http://dx.doi.org/10.1038/nrg2626

Nagele, P. (2011). Perioperative genomics. Best. Pract. Res. Clin. Anaesthesiol. *25*, 549–555. http://dx.doi.org/10.1016/j.bpa.2011.09.001

Nakamura, S., Maeda, N., Miron, I.M., Yoh, M., Izutsu, K., Kataoka, C., Honda, T., Yasunaga, T., Nakaya, T., Kawai, J., *et al.* (2008). Metagenomic diagnosis of bacterial infections. Emerg. Infect. Dis. *14*, 1784–1786. http://dx.doi.org/10.3201/eid1411.080589

Nassirpour, R., Mathur, S., Gosink, M.M., Li, Y., Shoieb, A.M., Wood, J., O'Neil, S.P., Homer, B.L., and Whiteley, L.O. (2014). Identification of tubular injury microRNA biomarkers in urine: comparison of next-generation sequencing and qPCR-based profiling platforms. BMC Genomics *15*,485.

Pabinger, S., Rödiger, S., Kriegner, A., Vierlinger, K., and Weinhäusel, A. (2014). A survey of tools for the analysis of quantitative PCR (qPCR) data. Biomol. Detect. Quantif. *1*, 23-33.

Peng, X., Yu, K.Q., Deng, G.H., Jiang, Y.X., Wang, Y., Zhang, G.X., and Zhou, H.W. (2013). Comparison of direct boiling method with commercial kits for extracting fecal microbiome DNA by Illumina sequencing of 16S rRNA tags. J. Microbiol. Methods *95*, 455–462. http://dx.doi.org/10.1016/j.mimet.2013.07.015

Perandin, F., Manca, N., Calderaro, A., Piccolo, G., Galati, L., Ricci, L., Medici, M.C., Arcangeletti, M.C., Snounou, G., Dettori, G., *et al.* (2004). Development of a real-time PCR assay for detection of Plasmodium falciparum, Plasmodium vivax, and Plasmodium ovale for routine clinical diagnosis. J. Clin. Microbiol. *42*, 1214–1219.

Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., and Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics *13*, 341.

Renkema, K.Y., Stokman, M.F., Giles, R.H., and Knoers, N.V. (2014). Next-generation sequencing for research and diagnostics in kidney disease. Nat. Rev. Nephrol. *10*, 433–444. http://dx.doi.org/10.1038/nrneph.2014.95

Ribao, C., Torrado, I., Vilariño, M.L., and Romalde, J.L. (2004). Assessment of different commercial RNA-extraction and RT-PCR kits for detection of hepatitis A virus in mussel tissues. J. Virol. Methods *115*, 177–182.

Schmittgen, T.D., Lee, E.J., Jiang, J., Sarkar, A., Yang, L., Elton, T.S., and Chen, C. (2008). Real-time PCR quantification of precursor and mature microRNA. Methods *44*, 31–38.

Schröder, K., and Nitsche, A. (2010). Multicolour, multiplex real-time PCR assay for the detection of human-pathogenic poxviruses. Mol. Cell. Probes *24*, 110–113. http://dx.doi.org/10.1016/j.mcp.2009.10.008

Schulze, M., Nitsche, A., Schweiger, B., and Biere, B. (2010). Diagnostic approach for the differentiation of the pandemic influenza A(H1N1)v virus from recent human influenza viruses by real-time PCR. PLOS ONE *5*, e9966. http://dx.doi.org/10.1371/journal.pone.0009966

Templeton, K.E., Scheltinga, S.A., Beersma, M.F., Kroes, A.C., and Claas, E.C. (2004). Rapid and sensitive method using multiplex real-time PCR for diagnosis of infections by influenza A and influenza B viruses, respiratory syncytial virus, and parainfluenza viruses 1, 2, 3, and 4. J. Clin. Microbiol. *42*, 1564-1569.

Voelkerding, K.V., Dames, S.A., and Durtschi, J.D. (2009). Next-generation sequencing: from basic research to diagnostics. Clin. Chem. *55*, 641–658. http://dx.doi.org/10.1373/clinchem.2008.112789

Yamamoto, Y. (2002). PCR in diagnosis of infection: detection of bacteria in cerebrospinal fluids. Clin. Diagn. Lab. Immunol. *9*, 508–514.