SIMPLIFIED HOST DNA REMOVAL PROCEDURE FOR VIRAL DETECTION IN CLINICAL BLOOD SAMPLES

Patsaporn Jarupathirun[1,2], Angkana Huang[3], Kwanrutai Chin-inmanu[2], Harald Grove[2], Piyawan Chinnawirotpisan[3], Thipwipha Phonpakobsin[3], In-Kyu Yoon[5], Stefan Fernandez[3], Louis R Macareo[3], Prapat Suriyaphol[2] and Wiriya Rutvisuttinunt[3,4]

[1]Molecular Medicine Graduate Program, Faculty of Science, Mahidol University; [2]Division of Bioinformatics and Data Management for Research, Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University; [3]Department of Virology, Armed Forces Research Institute of Medical Sciences (AFRIMS), Bangkok, Thailand; [4]Viral Diseases Branch, Walter Reed Army Institute of Research, Maryland, USA; [5]International Vaccine Institute, Seoul, Korea

Abstract. Identification of pathogenic viruses with accuracy and speed in clinical samples for diagnosis is crucial to prevent a disease outbreak. In a typical blood sample, the amount of host nucleic acids in vast excess compared to pathogen nucleic acids makes identification of an unknown pathogen extremely difficult and time consuming. In this study, we investigated the efficiency of DNase I and Omnicleave nucleases together with different centrifugation speeds to remove host DNA in plasma and serum samples spiked with dengue virus type-1 (DENV-1). The quantities of DENV-1 RNA and contaminating human DNA were evaluated by qRT-PCR and Qubit® fluorometer, respectively, to determine the most efficient procedure for host nucleic acid removal. Enzymatic digestion was inefficient in removing host DNA from plasma and serum samples, similarly, using low-speed (6,200*g*) centrifugation by itself, with 10% average reduction. The most effective procedure was a combination of low- and high-speed (23,500*g*) centrifugation steps, achieving 80% reduction in host DNA for both DENV-1 spiked plasma and serum samples. The procedure is rapid, simple to perform and can be easily incorporated into any DNA extraction workflow.

Keywords: blood sample, dengue virus, host DNA removal, next generation sequencing, viral enrichment

Correspondence: Prapat Suriyaphol, Division of Bioinformatics and Data Management for Research, Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, 2 Wang Lang Road, Bangkok Noi, Bangkok 10700, Thailand.
E-mail: prapat.sur@mahidol.ac.th
Wiriya Rutvisuttinunt, Department of Virology, Armed Forces Research Institute of Medical Sciences (AFRIMS), 315/6, Ratchawithi Road, Thung Phaya Thai, Ratchathewi, Bangkok 10400, Thailand.
E-mail: wiriya.rutvisuttinunt.ctr@mail.mil

INTRODUCTION

Infectious diseases are the second most common cause of death worldwide (Mathers *et al*, 2008). Infections from viral pathogens, such as dengue virus (Bhatt *et al*, 2013) and Ebola virus (Piot, 2016), pose a major threat to global public health. Early and accurate diagnosis can have a significant and positive impact on morbidity and mortality of disease outbreaks. Diagnosis of viral infections is a slow

and complex process because different pathogens may lead to similar clinical symptoms. Due to the limitations of viral detection methods, patients are treated based on their symptoms, while the etiology of their disease remains unknown.

Approximately 58% of febrile illnesses are caused by blood-borne viruses (Capeding *et al*, 2013). In addition to known viruses, there are also novel viruses, thereby increasing the importance of developing new detection methods for viral surveillance and discovery. New human viruses have recently been discovered by molecular analysis of samples from patients with symptoms of unknown etiology, *via* encephalitis (Glaser *et al*, 2003), hepatitis (Desai *et al*, 1999; He *et al*, 2003; El Gaafary *et al*, 2005), gastrointestinal diseases, myocarditis, and auto-immune disease (Hajjeh *et al*, 2002).

In recent years, next-generation sequencing (NGS) technologies have been improved, being able to generate many millions of sequences in a single round of sequencing (Mardis, 2008). However, they differ as regards cost, turnaround time, sequence yield per run, accuracy, error rate and read length (Metzker, 2010; Quail *et al*, 2012). One of the crucial steps in viral detection and identification when using NGS technologies is sample preparation. The quantity of viral genomic material is low in clinical specimens and the nucleic acid extraction and sample processing become crucial to obtain pathogen nucleic acids in good quality and quantity for detection.

Viral RNA is usually extracted from clinical samples and then reverse-transcribed into DNA (Datta *et al*, 2015). Although there are several methods for RNA isolation from clinical samples, *viz*. organic solution -based extraction, silica-membrane-based spin column and magnetic bead- based technology (Rump *et al*, 2010; Tavares *et al*, 2011; Sellin *et al*, 2014), one major problem with these isolation methods is that they result in high amounts of contaminating host DNA, removal of which will help increase the relative amount of viral RNA. However, there is a risk of losing some amount of viral RNA through the various removal procedures (Marston *et al*, 2013), causing a lower output of genetic sequence information or number of NGS reads (Delwart, 2007). In order to obtain sufficient genetic information from pathogen nucleic acids, many laboratories attempt repeated sequencing or culturing to increase the amount of nucleic acids required to the sequence reads (Coffey *et al*, 2014). These strategies increase cost and complicate results. Furthermore, the large amount of host nucleic acid in clinical specimens requires more pre-processing, filtration and other reduction of their signals in bioinformatics analysis to detect the pathogen nucleic acid sequence. These additional steps make detection and identification of the pathogen more difficult, slowing down the clinical diagnostic process (Radford *et al*, 2012).

Often physical methods or endonucleases are employed to remove host genomic DNA and enhance viral enrichment (Malboeuf *et al*, 2013; Manso *et al*, 2017). In this study, we evaluated the effectiveness of several methods mentioned in the literature used for host nucleic acid removal to improve viral enrichment and subsequent viral identification in spiked plasma and serum samples. The goal was to determine the most efficient, cost-effective and proficient procedure that can be applied to routine clinical diagnosis.

MATERIALS AND METHODS

Materials

Commercially processed plasma and pooled serum from healthy individuals were obtained from The Thai Red Cross Society. Both plasma and serum samples were confirmed negative for Japanese encephalitis virus and dengue virus (DENV) by hemagglutination inhibition (HAI) assays (Clarke and Casals, 1955). These serum and plasma samples are referred to as negative samples. Human genomic DNA for control samples was acquired from Jena Bioscience (Jena, Thuringia, Germany). DENV-1 Hawaii strain (GenBank accession number EU848545) was cultured in a C6/36 cell line as previously described (Jarman *et al*, 2011), and supernatant from the cell culture was stored at -70°C until used. RNase-free DNase I and Omnicleave™ endonuclease was from Ambion (Life Technologies, Auckland, New Zealand) and Epicentre (Madison, WI), respectively. DENV-1 was spiked in serum and plasma samples at physiological condition.

Nucleic acid extraction and measurement

For plasma and serum samples, (spiked) viral RNA was extracted using QIAamp® viral RNA Mini Kit (QIAGEN, Valencia, CA). Quantity of host DNA was measured pre- and post-treatments using Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA). The amount of DENV-1 in plasma and serum was measured using quantitative (q) RT-PCR as previously described (Sadon *et al*, 2008) with modifications. In brief, qRT-PCR was conducted using AgPath-ID™ One-Step RT-PCR Kit (Ambion, Life Technologies, Auckland, New Zealand) in a reaction volume of 22.5 µl. Thermocycling was performed in an ABI 7300 Real Time PCR system (Applied Biosystems, Foster City, CA) as follows: 45°C for 10 minutes; 95°C for 10 minutes: and 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. Values are expressed in genome equivalent/ml (GE/ml). DENV-1 spiked sample with Ct value of 23.8 was used for all experiments.

Nuclease digestion

Different parameters for nuclease digestion, including endonuclease concentration, Mg^{2+} concentration, temperature, and incubation time were assessed to refine the condition that yielded the lowest amount of host genomic material. All samples were centrifuged at low speed (see below) to remove cell debris. A total of 81 conditions were tested in combination of the following: (i) DNase I (1.5, 3 and 6 U), (ii) free Mg^{2+} concentration (0.5, 1 and 2 mM), (iii) temperature of incubation (22, 30 and 37 °C), and (iv) reaction time (15, 30 and 60 minutes).

Centrifugation

Low-speed centrifugation (6,200*g* at 4°C for 10 minutes) was used to sediment cell debris (Krejbich-Trotot *et al*, 2011). High-speed centrifugation (23,500*g* at 4°C for 20 minutes) was performed using a SS-34 Sorvall rotor (Thermo Fisher Scientific, Waltham, MA) (Cepko, 2001) following the low-speed centrifugation and the pellet was re-suspended with Gibco® RPMI 1640 medium (Invitrogen, Grand Island, NY) prior to RNA extraction.

Next-generation sequencing (NGS)

Extracted viral RNA was reverse transcribed to cDNA using SuperScript III (Life technologies, Invitrogen, Grand Island, NY). The double-stranded DNA was purified using AMPure XP beads (Beckman Coulter, Brea, CA). Libraries were constructed from DNA (1 ng) using an Illumina Nextera DNA sample preparation kit (Illumina, San Diego, CA) and sequenced using an Illumina

MiSeq instrument for 300 cycles. The sample preparation process was adapted from Illumina protocol and performed as previously described (Rutvisuttinunt *et al*, 2014).

Evaluation of treatment combinations

Sample treatments were evaluated based on their efficiency in reducing the amount of host DNA and maximizing the recovery of viral RNA. Using the untreated sample as control, percent remaining host DNA and viral RNA were used to select the optimal treatment conditions for treatment of serum and plasma samples. The formulas used for the calculations are as follows: (i) remaining host DNA = host DNA of treated sample/host DNA of untreated sample, (ii) recovered viral RNA = viral genome copies of treated sample/ viral genome copies of untreated sample, and (iii) ratio of signal retaining to noise retained = recovered viral RNA/remaining host DNA. Host DNA in the formulas refers to the nucleic acid quantity measured in ng/μl by Qubit® fluorometer. Viral genome copies were calculated from the Ct values obtained from the qRT-PCR assay. Statistical significance is calculated using pairwise *t*-test in R (v. 3.1.1) (R Core Team, 2016). A p-value <0.05 is considered significant.

Bioinformatics analysis

Raw sequenced reads were filtered using PRINSEQ software version 0.20.4 (Schmieder and Edwards, 2011) to exclude reads shorter than 40 bases, with an average quality score <25, containing more than one ambiguous base call (Ns), or having an exact duplicate with 100% sequence identity. Cleaned sequence reads were then mapped to the two selected reference genomes, namely, human (hg38, NCBI) and DENV-1 isolate US/Hawaii/1944 (GenBank EU848545). Reads were first mapped to the human reference, and those not finding a hit were then mapped to the DENV reference. Two alignment methods were used, namely, Bowtie2 ultrafast sequence aligner with default parameters (Langmead and Salzberg, 2012) and Basic Local Alignment Search Tool (BLAST). Results from the BLAST alignment were reported as the number of reads mapped to the human genome with 100% sequence coverage and to the DENV-1 with at least 80% sequence coverage. Recovery of virus after treatments was assessed by comparing the number of mapped reads between treated and untreated samples.

The overall workflow is shown in Fig 1.

RESULTS

Nuclease digestion

We first evaluated one of the most commonly used nucleases for host removal procedure, DNase I, an endonuclease that nonspecifically cleaves single- and double-stranded (ss and ds) DNA, chromatin and RNA:DNA hybrids (Silha and Kolibova, 1980; Stewart *et al*, 1991; Suck, 1994; Akaboshi, 1999). Evaluation of the effects of different parameters (temperature, Mg^{2+} concentration, enzyme concentration, and digestion time), shown to be critical for enzyme function and often vary digestion outcome, revealed the largest reduction, 7%, occurred under the following conditions: DNase I = 1.5 U, free Mg^2 concentration = 1 mM, temperature = 37°C, and time = 15 minutes (Fig 2). We also observed that 70% (57/81) of the tested parameter combinations resulted in an increase in measurable DNA content.

Centrifugation with and without prior nuclease digestion

We evaluated both low- and high-speed centrifugations with and without

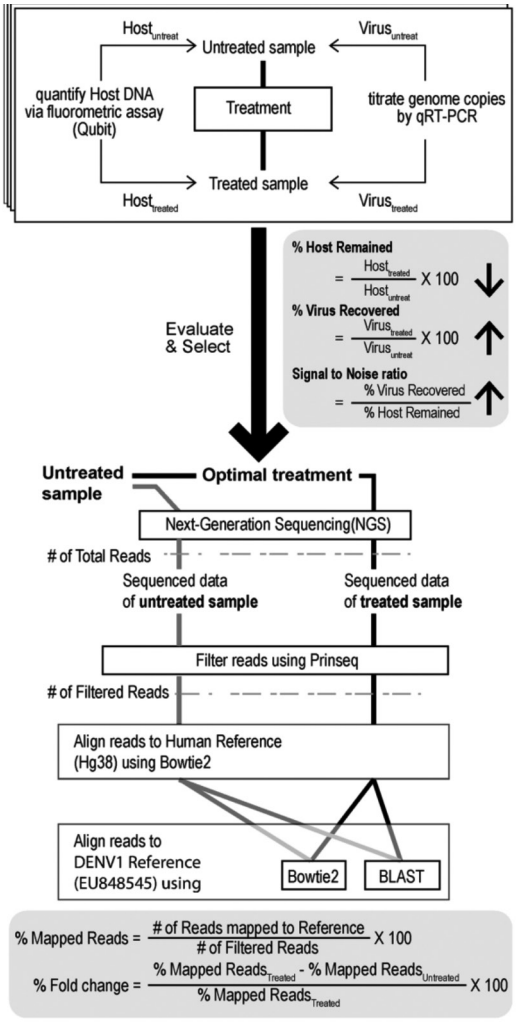

Fig 1–Flow diagram of study procedures and formulas for result assessments.

prior nuclease digestion on the removal of host DNA. Due to the differences in the specificity of various nucleases, the additional nuclease that can digest ss-DNA, dsDNA and RNA (Omnicleave™ endonuclease) was also used in the experiment to determine whether it could improve viral yield. In order to investigate the effect of low-speed centrifugation and to test whether enzyme treatment under the optimized condition from previous iteration (1.5 unit of enzyme DNaseI and 1 mM of magnesium concentration for 15 minutes at 37°C) is suitable for both enzymes (DNaseI and Omnicleave) or not, the comparison between the sample processing by low-speed centrifugation only and low-speed centrifugation combined with two different enzymes treatment (DNaseI and Omnicleave) at two different temperatures (37°C or 22°C) was demonstrated in plasma and serum samples. Centrifugation at low-speed alone reduced host DNA comparable to that of nuclease (DNase I or Omnicleave) digestion alone (Table 1); host DNA reduction using only Omnicleave compared to only low centrifugation force was better with serum (16% *vs* 4%) than plasma (9% *vs* 11%) sample. However, combining the two treatments did not improve the efficiency of host DNA removal (Table 1). Host DNA removal between combination of high speed centrifugation with (80%) and without (10%) nuclease digestion is statistical significant (p <0.001 and 0.002 for plasma and serum sample, respectively) (Table 1). The amount of recovered viral RNA is not significantly different between low- and high-speed centrifugation. All combinations without high speed centrifugation had a signal to noise ratio of <1. The highest ratio was obtained with low- followed by high-speed centrifugation for plasma sample or with Omnicleave digestion and sequential low- and high-speed centrifugation for serum sample (Table 1). In order to reduce the number of steps, prior nuclease digestion was omitted from the final procedure for both plasma and serum samples (Table 1). The several processes were combined. Different speeds of centrifugation are used in molecular laboratory procedures: low speed centrifugation for removing cell debris and high speed for pelleting

15 min	22°C	101%	94%	97%	106%	108%	105%	106%	99%	100%
	30°C	99%	97%	99%	104%	105%	106%	108%	99%	96%
	37°C	104%	93%	99%	104%	102%	100%	104%	96%	97%
30 min	22°C	107%	97%	101%	102%	105%	105%	113%	99%	97%
	30°C	101%	95%	98%	103%	98%	97%	109%	99%	98%
	37°C	104%	98%	98%	101%	101%	102%	106%	103%	99%
60 min	22°C	108%	103%	103%	106%	111%	108%	113%	103%	106%
	30°C	110%	107%	108%	115%	112%	112%	116%	104%	104%
	37°C	108%	109%	106%	107%	120%	109%	120%	104%	110%
Time	Temp									
Mg (mM)		0.5	1	2	0.5	1	2	0.5	1	2
DNAsel (U)		1.5			3			6		

Fig 2–DNase I efficiency in digesting human DNA under various conditions. Percent of remaining host DNA in the sample was measured and calculated after each treatment configuration with varying incubation time (15 min, 30 min, 60 min), temperature (22ºC, 30ºC, 37ºC), magnesium concentration (0.5 mM, 1 mM, 2 mM) and unit of DNAseI (1.5 U, 3 U, 6 U).

viral particles. We evaluated both high speed and low speed centrifugation with and without enzymatic digestion and accessed the removal of host DNA (Treatment 1, Table 1, section 1). The proportion of recovered viruses was measured and used in the evaluation for viral genome recovery (Treatment 2, Table 1, section 2). Optimal treatments were performed in triplicate (Treatment 3, Table 1, section 3) to determine the effectiveness of enriching for viral RNA using low-speed centrifugation followed by high-speed centrifugation. The fold ratio was calculated after low-speed centrifugation (3B), followed by Omnicleave (3C, 3D) and after high-speed centrifugation was completed (3E, 3F). One condition used only low-speed centrifugation followed by high-speed centrifugation omitting endonuclease treatment (3G).

Effect of virus enrichment on sequence reads

A new set of DENV-1 spiked samples were prepared and treated with low-speed and high-speed centrifugation [Table 2: Treatment A (control) and B]. Based on previous experience (unpublished data), an additional sample that was prepared using a longer high-speed centrifugation time (60 minutes) was also included (Table 2: Treatment C and C-2). All samples were sequenced using NGS, which demonstrated an improvement in the ratio of virus sequence to human sequence reads for each treatment (Table 2). The results of the quantitative RT-PCR corresponded with the NGS data (data not shown). All treatments increased the ratio of viral reads, but the effect was 5-7 times higher for plasma than serum samples. There was also a 4.2- and 5.6-fold increase between performing high-speed centrifugation for 60 than 20 minutes for plasma and serum sample, respectively.

The number of reads mapping to the human and the DENV-1 genomes, after Q30 quality trimming, varied between treatments due to experimental differences including the number of samples analyzed together (Table 2). Overall, Treat-

Table 1
Recovery of host DNA and viral RNA from treated plasma and serum samples.

Treatment		Endonuclease digestion				Centrifugation		Outcome					
		DNase I		Omnicleave				Plasma sample			Serum sample		
		37°C, 15 min	22°C, 15 min	37°C, 15 min	22°C, 15 min	Low-speed[a]	High-speed[b]	Percent remaining host DNA	Percent virus RNA recovered	Signal: noise	Percent remaining host DNA	Percent virus RNA recovered	Signal: noise
1	A[c]							100	-	-	100	-	-
	B	x						96	-	-	97	-	-
	C		x					101	-	-	95	-	-
	D			x				91	-	-	83	-	-
	E				x			91	-	-	84	-	-
	F					x		89 ± 1[d]	-	-	95 ± 4[d]	-	-
	G	x				x		96	-	-	105	-	-
	H		x			x		90	-	-	94	-	-
	I			x		x		91	-	-	85	-	-
	J				x	x		90	-	-	84	-	-
2	A[c]							100	100	1.00	100	100	1.00
	B			x				91	36	0.40	88	40	0.45
	C				x			91	69	0.76	91	68	0.75
	D					x		88	28	0.32	-	-	-
	E				x	x		-	-	-	88	10	0.11
	F			x			x	21	27	1.29	21	79	3.76
	G				x		x	24	45	1.88	14	69	4.93
	H					x	x	25	79	3.16	-	-	-
	I				x	x	x	-	-	-	15	139	9.27
3	A[c]							100	100	1	100	100	1.00
	B					x		96 ± 2[e]	92	0.96	89 ± 1[e]	101	1.13
	C			x		x		93 ± 4[e]	93	1	-	-	-
	D				x	x		-	-	-	85 ± 2[e]	113	1.33
	E			x		x	x	16 ± 2[e]	17	1.09	-	-	-
	F				x	x	x	-	-	-	10.0 ± 0.1[e]	22	2.20
	G					x	x	14 ± 1[e]	18	1.25	10.6 ± 0.6[e]	70	6.60

[a]6,200*g*, 4°C, 10 min. [b]23,500*g*, 4°C, 20 min. [c]No treatment; used as baseline quantity for percent calculation. [d]Mean ± SD; quadruplet experiments. [e]Mean ± SD; triplicate experiments. Min, minutes; Signal:noise, percent virus RNA recovered / percent remaining host DNA.

Table 2
Bioinformatics results for plasma and serum samples.

Process	Bowtie 2		BLAST
	Human (Hg38)	Dengue virus 1 (EU848545)	Dengue virus 1 (EU848545)
Plasma samples			
Treatment A (total reads = 421,242; filtered reads[a] =359,524)			
Number of reads	334,658	1,667	1,822
(%)	(93.08)	(0.46)	(0.51)
Fold change (%)	0	0	0
Treatment B (total reads = 39,224; filtered reads[a] = 31,798)			
Number of reads	18,443	516	527
(%)	(58)	(1.62)	(1.66)
Fold change (%)	-37.69	249.98	227.03
Treatment C (total reads = 79,786; filtered reads[a] = 63,216)			
Number of reads	40,651	4,748	4,749
(%)	(64.3)	(7.51)	(7.51)
Fold change (%)	-30.92	1,519.85	1,382.36
Treatment C2 (verification of Treatment C) (total reads = 52,722; filtered reads[a] = 43,207)			
Number of reads	27,474	3238	3207
(%)	(63.59)	(7.49)	(7.42)
Fold change (%)	-31.69	1,516.27	1,364.62
Serum samples			
Treatment A (total reads = 2,367,498; filtered reads[a] = 2,045,876)			
Number of reads	1,923,671	1,510	1,665
(%)	(94.03)	(0.07)	(0.08)
Fold change (%)	0	0	0
Treatment B (total reads = 715,354; filtered reads[a] = 624,501)			
Number of reads	565,089	350	527
(%)	(90.49)	(0.06)	(0.08)
Fold change (%)	-3.77	(24.07	3.69
Treatment C (total reads = 1,310,950; filtered reads[a] = 1,134,592)			
Number of reads	1,054,465	3,251	3,260
(%)	(92.94)	(0.29)	(0.29)
Fold change (%)	-1.16	288.22	253.06
Treatment C2 (verification of Treatment C) (total reads = 226,770; filtered reads[a] = 199,492)			
Number of reads	184,772	722	712
(%)	(92.62)	(0.36)	(0.36)
Fold change (%)	-1.50	390.36	338.55

[a]Number of reads that passed data filtration step using PRINSEQ. Treatment A, (No treatment); Treatment B, (Low-speed centrifugation followed by high-speed centrifugation for 20 min); Treatment C, (Low-speed centrifugation followed by high-speed centrifugation for 1 hr).

ment C showed a consistent proportion of virus to human reads. We also observed an increase in unmapped reads after treatment for both plasma (4.4-6.3 folds) and serum (1.2-1.6 folds). When these unmapped reads were analyzed separately using BLAST they were found to map to the human reference genome, but below the set limit for counting as a valid hit.

DISCUSSION

The aim of this study was to optimize methods for removing host nucleic acid from blood samples to increase the relative amount of pathogen nucleic acid to host genome thereby increasing the probability of pathogen identification. Performing high-speed centrifugation (23,500*g*) following endonuclease digestion or low-speed (6,200*g*) followed by high-speed centrifugation was the optimal protocol to remove (80%) host DNA, with better results using infected serum than plasma samples.

Although Omnicleave performed better than DNase I, neither enzyme alone reduced host DNA by more than 20%. DNase I is commonly used for host nucleic acids removal to increase signal of viral pathogens (Allander *et al*, 2001; Rutvisuttinunt *et al*, 2014; *ibid*, 2015; Matranga *et al*, 2016). Although higher amounts of endonuclease is believed to be more efficient (Robinson, 2015), our results showed optimal digestion was achieved with the minimal amount of enzyme. Surprisingly, under certain digestion conditions the amount of DNA measured by Qubit® fluorometer increased after incubation, possibly due to other degradation processes that occurred and thereby releasing DNA fragments. One possible source of such DNA might be extracellular vesicles, microparticles, exosomes or apoptotic bodies released from cells, in addition to other molecules from the originating cell (El Andaloussi *et al*, 2013; Thakur *et al*, 2014; Zaborowski *et al*, 2015). During incubation, these vesicles could release nucleotides and this combined with a reduced effectiveness of the endonuclease under these conditions might explain the phenomenon. Furthermore, DNA might also be unavailable for measurement due to being bound to proteins, in which case the increased temperature during incubation could release DNA.

The low efficiency of DNase I in removing host DNA is in contradiction to work by Hall *et al* (2014) investigating DNA removal samples containing A549 human epithelial lung carcinoma cells spiked with *Escherichia coli* O157, human enterovirus 71, human adenovirus 5 and influenza A (H1N1) pdm09. This could be accounted for by the difference in the samples investigated.

For sequence analysis, we found BLAST to be more flexible than Bowtie 2 when the numbers of sequence reads were low. The two different mapping techniques performed equally well with BLAST mapping slightly more reads. Bowtie2 includes read quality in determining a valid alignment, but is faster and more useful than BLAST when time is limited.

The relative amount of viral sequence reads from plasma was significantly higher when compared to serum despite all samples having been spiked with the same amount of virus particles. It is possible that the centrifugation steps were affected by differences in the composition and/or physical properties of the two types of samples.

In conclusion, enrichment of virus RNA in blood samples required methods optimized for specific sample types. Enzy-

matic removal of host DNA did not provide additional benefit to the sequential low- and high-speed centrifugation steps demonstrated to provide a robust enrichment procedure for viral RNA.

ACKNOWLEDGEMENTS

The authors thank Dr Butsaya Thaisomboonsuk, Dr Chuanpis Ajariyakhajorn, Ms Duangrat Mongkolsirichaikul, Mr Winai Kaneechit, Ms Prapapun Ongajchaowlerd, and the virology and serology laboratory team for processing viral isolates and commercial processed plasma samples; Dr Charlotte Lanteri and the Department of Immunology, Faculty of Medicine Siriraj Hospital, Mahidol University for commercial processed serum; the clinical coordinating research team and project management section for specimen processing and demographic data on study subjects; Dr Chonticha Klungthong, Ms Prinyada Rodpradit, Mr Kittinun Hussem and the molecular section team for laboratory technical assistance with the next generation sequencing, Qubit® fluorometer and bioanalyzer; and Mr Adison Khongchatee for assistance in preparing the manuscript.

Funding of this work was supported by the Armed Forces Health Surveillance Center-Global Emerging Infections Surveillance and Response System (AFHSC-GEIS). P Jarupathirun was supported by the scholarship from Medical Scholars Program of Mahidol University. P Suriyaphol was granted by TRF-Office for R&D (grant number IRG5980006).

The views and opinions expressed herein are the private opinions of the authors and do not necessarily reflect those of the US Army or the Department of Defense.

REFERENCES

Allander T, Emerson SU, Engle RE, Purcell RH, Bukh J. A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proc Natl Acad Sci USA* 2001; 98: 11609-14.

Akaboshi, E. Recognition of sequence-directed structure of the ssDNA backbone by nucleases. *Biochem Biophys Res Commun* 1999; 256: 532-6.

Bhatt S, Gething PW, Brady OJ, *et al*. The global distribution and burden of dengue. *Nature* 2013; 496: 504-7.

Capeding MR, Chua MN, Hadinegoro SR, *et al*. Dengue and other common causes of acute febrile illness in Asia: an active surveillance study in children. *PLOS Negl Trop Dis* 2013; 7: e2331.

Clarke DH, Casals J. Improved methods for hemagglutination studies with arthropodborne viruses. *Proc Soc Exp Biol Med* 1955; 88: 96-9.

Cepko C. Large-scale preparation and concentration of retrovirus stocks. *Curr Protoc Mol Biol* 2001; Chapter 9: Unit9. 12.

Coffey LL, Page BL, Greninger AL, *et al*. Enhanced arbovirus surveillance with deep sequencing: identification of novel rhabdoviruses and bunyaviruses in Australian mosquitoes. *Virology* 2014; 448: 146-58.

Datta S, Budhauliya R, Das B, Chatterjee S, Vanlalhmuaka, Veer V. Next-generation sequencing in clinical virology: discovery of new viruses. *World J Virol* 2015; 4: 265-76.

Delwart EL. Viral metagenomics. *Rev Med Virol* 2007; 17: 115-31.

Desai SM, Muerhoff AS, Leary TP, *et al*. Prevalence of TT virus infection in US blood donors and populations at risk for acquiring parenterally transmitted viruses. *J Infect Dis* 1999; 179: 1242-4.

El Andaloussi S, Mager I, Breakefield XO, Wood MJ. Extracellular vesicles: biology and

emerging therapeutic opportunities. *Nat Rev Drug Discov* 2013; 12: 347-57.

El Gaafary MM, Rekacewicz C, Abdel-Rahman AG, *et al*. Surveillance of acute hepatitis C in Cairo, Egypt. *J Med Virol* 2005; 76: 520-5.

Glaser CA, Gilliam S, Schnurr D, *et al*. In search of encephalitis etiologies: diagnostic challenges in the California Encephalitis Project, 1998-2000. *Clin Infect Dis* 2003; 36: 731-42.

Hajjeh RA, Relman D, Cieslak PR, *et al*. Surveillance for unexplained deaths and critical illnesses due to possibly infectious causes, United States, 1995-1998. *Emerg Infect Dis* 2002; 8: 145-53.

Hall RJ, Wang J, Todd AK, *et al*. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *J Virol Methods* 2014; 195: 194-204.

He Z, Zhuang H, Wang X, *et al*. Retrospective analysis of non-A-E hepatitis: possible role of hepatitis B and C virus infection. *J Med Virol* 2003; 69: 59-65.

Jarman RG, Nisalak A, Anderson KB, *et al*. Factors influencing dengue virus isolation by C6/36 cell culture and mosquito inoculation of nested PCR-positive clinical samples. *Am J Trop Med Hyg* 2011; 84: 218-23.

Krejbich-Trotot P, Denizot M, Hoarau JJ, Jaffar-Bandjee MC, Das T, Gasque P. Chikungunya virus mobilizes the apoptotic machinery to invade host cell defenses. *FASEB J* 2011; 25: 314-25.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012; 9: 357-9.

Malboeuf CM, Yang X, Charlebois P, *et al*. Complete viral RNA genome sequencing of ultra-low copy samples by sequence-independent amplification. *Nucleic Acids Res* 2013; 41: e13.

Manso CF, Bibby DF, Mbisa JL. Efficient and unbiased metagenomic recovery of RNA virus genomes from human plasma samples. *Sci Rep* 2017; 7: 4173.

Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 2008; 9: 387-402.

Marston DA, McElhinney LM, Ellis RJ, *et al*. Next generation sequencing of viral RNA genomes. *BMC Genomics* 2013; 14: 444.

Mathers C, Boerma T, Fat DM. The global burden of disease : 2004 update. Geneva: World Health Organization, 2008.

Matranga CB, Gladden-Young A, Qu J, *et al*. Unbiased deep sequencing of RNA viruses from clinical samples. *J Vis Exp* 2016 Jul 2; (113).

Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet* 2010; 11: 31-46.

Piot P. Public health: beating Ebola. *Nature* 2016; 537: 484-5.

Quail MA, Smith M, Coupland P, *et al*. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 2012; 13: 341.

Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N. Application of next-generation sequencing technologies in virology. *J Gen Virol* 2012; 93: 1853-68.

R Core Team. A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2016.

Robinson PK. Enzymes: principles and biotechnological applications. *Essays Biochem* 2015; 59: 1-41.

Rump LV, Asamoah B, Gonzalez-Escalona N. Comparison of commercial RNA extraction kits for preparation of DNA-free total RNA from Salmonella cells. *BMC Res Notes* 2010; 3: 211.

Rutvisuttinunt W, Chinnawirotpisan P, Klungthong C, *et al*. Evidence of West Nile virus infection in Nepal. *BMC Infect Dis* 2014; 14: 606.

Rutvisuttinunt W, Chinnawirotpisan P, Thaisomboonsuk B, *et al*. Viral subpopulation diversity in influenza virus isolates com-

pared to clinical specimens. *J Clin Virol* 2015; 68: 16-23.

Sadon N, Delers A, Jarman RG, *et al*. A new quantitative RT-PCR method for sensitive detection of dengue virus in serum samples. *J Virol Methods* 2008; 153: 1-6.

Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 2011; 27: 863-4.

Sellin Jeffries MK, Kiss AJ, Smith AW, Oris JT. A comparison of commercially-available automated and manual extraction kits for the isolation of total RNA from small tissue samples. *BMC Biotechnol* 2014; 14: 94.

Silha M, Kolibova A. Solubilization of chromatin by means of DNase I. *Folia Biol (Praha)* 1980; 26: 116-22.

Stewart AF, Reik A, Schutz G. A simpler and better method to cleave chromatin with DNase 1 for hypersensitive site analyses. *Nucleic Acids Res* 1991; 19: 3157.

Suck D. DNA recognition by DNase I. *J Mol Recognit* 1994; 7: 65-70.

Tavares L, Alves PM, Ferreira RB, Santos CN. Comparison of different methods for DNA-free RNA isolation from SK-N-MC neuroblastoma. *BMC Res Notes* 2011; 4: 3.

Thakur BK, Zhang H, Becker A, *et al*. Double-stranded DNA in exosomes: a novel biomarker in cancer detection. *Cell Res* 2014; 24: 766-9.

Zaborowski MP, Balaj L, Breakefield XO, Lai CP. Extracellular vesicles: composition, biological relevance, and methods of study. *Bioscience* 2015; 65: 783-97.