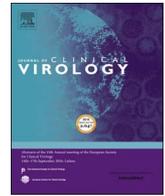




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Retrospective use of next-generation sequencing reveals the presence of Enteroviruses in acute influenza-like illness respiratory samples collected in South/South-East Asia during 2010–2013



Wiriya Rutvisuttinunt<sup>a,d,\*</sup>, Chonticha Klungthong<sup>a</sup>, Butsay Thaisomboonsuk<sup>a</sup>, Piyawan Chinnawirotpisan<sup>a</sup>, Chuanpis Ajariyakhajorn<sup>a</sup>, Wudtichai Manasatienkij<sup>a</sup>, Thipwipha Phonpakobsin<sup>a</sup>, Chanthap Lon<sup>b</sup>, David Saunders<sup>b</sup>, Sonam Wangchuk<sup>f</sup>, Sanjaya K. Shrestha<sup>d,e</sup>, John Mark S. Velasco<sup>a</sup>, Maria Theresa P. Alera<sup>a</sup>, Sriluck Simasathien<sup>c</sup>, Daranee Buddhari<sup>a</sup>, Richard G. Jarman<sup>a</sup>, Louis R Macareo<sup>a</sup>, In-Kyu Yoon<sup>a</sup>, Stefan Fernandez<sup>a,\*</sup>

<sup>a</sup> Department of Virology, Armed Forces Research Institute of Medical Sciences, 315/6, Rajavithi Road, Rajathewi, Bangkok, Thailand

<sup>b</sup> Department of Immunology and Medicine, Armed Forces Research Institute of Medical Sciences, 315/6, Rajavithi Road, Rajathewi, Bangkok, Thailand

<sup>c</sup> Phramongkutklo Hospital, 315, Rajavithi Road, Rajathewi, Bangkok, Thailand

<sup>d</sup> Walter Reed/AFRIMS Research Unit Nepal, Kathmandu, Nepal

<sup>e</sup> Center for International Health, University of Bergen, Norway

<sup>f</sup> Royal Centre for Disease Control, Department of Public Health, Ministry of Health, Thimphu, Bhutan

### ARTICLE INFO

#### Keywords:

Respiratory  
Enterovirus  
Surveillance  
Influenza-like illness  
Next-generation sequencing  
South East Asia  
South Asia

### ABSTRACT

**Background:** Emerging and re-emerging respiratory pathogens represent an increasing threat to public health. Etiological determination during outbreaks generally relies on clinical information, occasionally accompanied by traditional laboratory molecular or serological testing. Often, this limited testing leads to inconclusive findings.

The Armed Forces Research Institute of Medical Sciences (AFRIMS) collected 12,865 nasopharyngeal specimens from acute influenza-like illness (ILI) patients in five countries in South/South East Asia during 2010–2013. Three hundred and twenty-four samples which were found to be negative for influenza virus after screening with real-time RT-PCR and cell-based culture techniques demonstrated the potential for viral infection with evident cytopathic effect (CPE) in several cell lines.

**Objective:** To assess whether whole genome next-generation sequencing (WG-NGS) together with conventional molecular assays can be used to reveal the etiology of influenza negative, but CPE positive specimens.

**Study design:** The supernatant of these CPE positive cell cultures were grouped in 32 pools containing 2–26 supernatants per pool. Three WG-NGS runs were performed on these supernatant pools. Sequence reads were used to identify positive pools containing viral pathogens. Individual samples in the positive pools were confirmed by qRT-PCR, RT-PCR, PCR and Sanger sequencing from the CPE culture and original clinical specimens.

**Results:** WG-NGS was an effective way to expand pathogen identification in surveillance studies. This enabled the identification of a viral agent in 71.3% (231/324) of unidentified surveillance samples, including common respiratory pathogens (100/324; 30.9%): enterovirus (16/100; 16.0%), coxsackievirus (31/100; 31.0%), echovirus (22/100; 22.0%), human rhinovirus (3/100; 3%), enterovirus genus (2/100; 2.0%), influenza A (9/100; 9.0%), influenza B, (5/100; 5.0%), human parainfluenza (4/100; 4.0%), human adenovirus (3/100; 3.0%), human coronavirus (1/100; 1.0%), human metapneumovirus (2/100; 2.0%), and mumps virus (2/100; 2.0%), in addition to the non-respiratory pathogen herpes simplex virus type 1 (HSV-1) (172/324; 53.1%) and HSV-1 co-infection with respiratory viruses (41/324; 12.7%).

\* Corresponding authors at: Department of Virology, Armed Forces Research Institute of Medical Sciences, 315/6, Rajavithi Road, Rajathewi, Bangkok, 10400, Thailand.  
E-mail addresses: [wiriya.rutvisuttinunt.ctr@mail.mil](mailto:wiriya.rutvisuttinunt.ctr@mail.mil) (W. Rutvisuttinunt), [Stefan.fernandez.mil@mail.mil](mailto:Stefan.fernandez.mil@mail.mil) (S. Fernandez).

## 1. Background

Asia is an epicenter for the emergence of diverse respiratory pathogens representing major global public health threats [1–3]. Accurate molecular identification of respiratory pathogens is a vital decision-making tool for public health officials and health care providers. Unfortunately, traditional laboratory techniques do not allow a broad detection of viral pathogens during routine surveillance [4].

Next-generation sequencing (NGS) is a powerful tool to detect emerging or re-emerging pathogens, and to obtain information about frequency of intra-host genetic variation and virological responses to treatments [5–7]. NGS has been utilized to study commonly circulating respiratory viruses and novel viruses found in vectors, animals, and humans [8–10]. In addition, whole genome NGS (WG-NGS) has been utilized to identify viruses in diagnostic clinical virology [11–13] and occasionally applied in parallel with standard diagnostic assays [12,14].

During 2010–2013, the Armed Forces Research Institute of Medical Sciences (AFRIMS) conducted routine acute respiratory illness surveillance in 5 countries in South/South East (S/SE) Asia: Nepal, Bhutan, Thailand, Cambodia, and Philippines. 12,865 nasal and/or oropharyngeal swabs in universal transport media (UTM) were collected from individuals seeking medical care for acute influenza-like illness (ILI). These specimens were tested for influenza virus (IFV) by rRT-PCR, and 52.4% IFV negative specimens were sent for cell-based culture isolation. Of these, 324 produced CPE but were negative for common respiratory virus serology tests suggesting the presence of uncommon respiratory viruses in the culture. We sought to determine if we could identify a pathogen in CPE positive samples despite the samples testing negative using antibody panels for common respiratory samples. The information obtained will be used to improve current molecular detection protocols, which can then be incorporated in routine surveillance and clinical diagnosis.

## 2. Objectives

To investigate whether WG-NGS can be used to uncover the etiology of IFV negative, but CPE positive specimens, we utilized the Illumina MiSeq platform in conjunction with traditional molecular assays.

## 3. Study design

Viral isolations of IFV rRT-PCR negative that showed CPE positive were pooled and sequenced by WG-NGS. NGS sequences were utilized to identify positive pools containing viral pathogens. PCRs and Sanger sequencing were further performed on individuals of the positive pools to identify the presence of specific viral genotypes in the positive specimens.

## 4. Materials and methods

### 4.1. ILI surveillance

Nasal and/or oropharyngeal swabs in UTM (Copan Diagnostics Inc., California, USA) were collected from ILI surveillance sites in S/SE Asia during 2010–2013. The surveillance samples were collected in Nepal, Bhutan, Thailand, Cambodia and Philippines under protocols approved by the Institutional Review Boards of host country institutions and the Walter Reed Army Institute of Research (WRAIR).

### 4.2. Viral isolation, serological and molecular testing

Respiratory specimens negative by IFV rRT-PCR underwent viral isolation in Madin-Darby canine kidney (MDCK) and Hep-2 cells (Fig. 1) [15,16]. Culture supernatant from CPE positive cultures were tested by hemagglutination inhibition (HAI) for IFV and by indirect

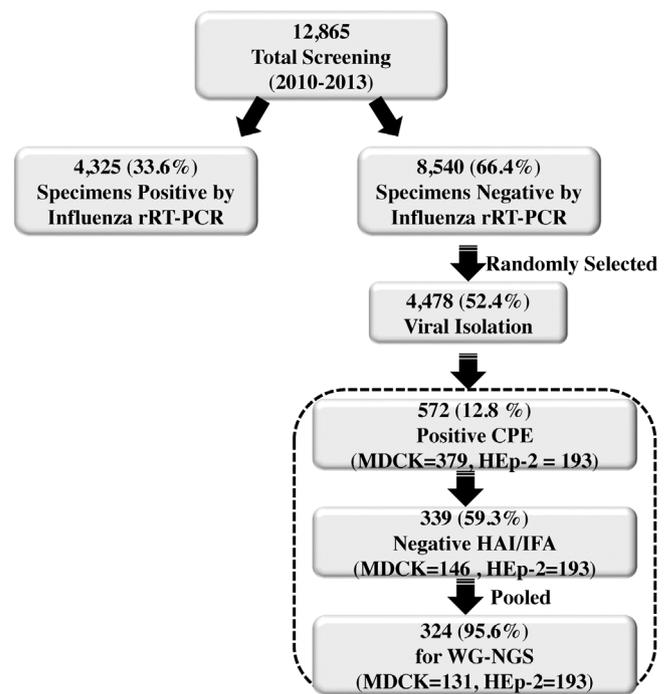


Fig. 1. Flow diagram illustrates the steps for influenza screening and identification of viral pathogens from respiratory specimens by whole genome next generation sequencing (WG-NGS).

fluorescent antibody (IFA) for respiratory syncytial virus (RSV), IFV A, IFV B, human adenovirus (HAdV), human parainfluenza (hPIV) 1, 2 and 3. CPE positive, but HAI and IFA-negative supernatants were selected for further testing by WG-NGS.

### 4.3. Pooled cultures and viral extraction

The supernatant of these CPE positives were grouped in 32 pools containing 2–26 supernatants per pool based on their country of origin, the year they were initially collected and cell line used for isolation (Tables 1A and 1B). Equal volume of each CPE positive supernatant was pooled to make up a total of volume of 300  $\mu$ l/pool. Virus enrichment steps, including removal of host cells and virus concentrating steps were applied as described previously [16,17]. QIAamp Viral RNA Extraction Kit (QIAGEN, CA, USA) was used for viral nucleic acid extraction. NanoDrop (Wilmington, DE, USA) and Agilent TapeStation1100 (Agilent Technologies, CA, USA) were utilized to quantify and determine the quality of the DNA and RNA (OD<sub>260</sub>/OD<sub>280</sub> values).

### 4.4. Sample preparation for WG-NGS

TruSeq LT RNA sample preparation was conducted following the manufacturer's instructions as described before [16,17]. Briefly, RNA was heat-fragmented, followed by double stranded cDNA synthesis (Life technologies, NY, USA) and adapter ligation. DNA quantity and quality control was validated and monitored by Agilent TapeStation1100, Qubit® 2.0 Fluorometer (Life Technologies, NY, USA) and rRT-PCR (Applied Biosystems, CA, USA). Three runs (10–12 pools/run) were conducted with Illumina MiSeq reagent version2 according to the manufacturer's protocol.

### 4.5. NGS data analysis

The quality control of sequence reads was done as previously described [16,17] with additional steps used for novel pathogen findings [Fig. 2]. Identification of individual viruses from the pool was performed as follows: Viral genome quantification was conducted by (i) de

**Table 1A**  
Viral pathogens identification of the pooled MDCK culture collected from South/South East Asia 2010–2013.

Pool No.	Year	Country	# Samples	# Viral Sequence Reads	Numbers of sequence reads identified as respiratory viruses (% identified sequence reads of identified virus/total viral sequence reads)											
					HSV-1	MuV	EV	CV	E	hRV	IFV A	IFV B	hPIVs	HAdV-1	HCoV	hMPV
1	2010, 2012	Nepal	8	71,502	71,143 (99.50%)	126* (0.28%)	95* (0.13%)	111* (0.16%)	-	-	-	27 (0.04%)	-	-	-	-
2	2013	Nepal	6	67,143	67,133 (99.99%)	-	-	-	-	-	-	10 (0.01%)	-	-	-	-
3	2010-2012	Bhutan	7	31,015	18,862 (60.82%)	120* (0.39%)	57* (0.18%)	108* (0.35%)	-	-	-	-	-	11,868 (38.27%)	-	-
4	2013	Bhutan	5	67,149	67,133 (99.8%)	-	-	-	-	-	-	16 (0.02%)	-	-	-	-
5	2010, 2012	Thailand	7	29,249	17,792 (60.83%)	79* (0.27%)	66* (0.23%)	18* (0.06%)	7* (0.02%)	-	-	11,287 (38.59%)	-	-	-	-
6	2010, 2012	Thailand	8	49,415	49,219 (99.60%)	102* (0.21%)	67* (0.14%)	24* (0.05%)	3* (0.01%)	-	-	-	-	-	-	-
7	2011	Thailand	15	46,742	46,673 (99.85%)	51* (0.11%)	15* (0.03%)	3* (0.01%)	-	-	-	-	-	-	-	-
8	2013	Thailand	10	263,342	263,335 (100.00%)	-	-	7* (0.00%)	-	-	-	-	-	-	-	-
9	2013	Thailand	10	181,729	181,729 (100.00%)	-	-	-	-	-	-	-	-	-	-	-
10	2013	Thailand	15	6,5714	65,005 (98.92%)	-	-	-	-	-	-	709 (1.08%)	-	-	-	-
11	2010-2012	Cambodia	11	634,579	49,743 (7.84%)	584,732 (92.14%)	45* (0.01%)	38* (0.01%)	17* (0.00%)	-	4 (0.00%)	-	-	-	-	-
12	2013	Cambodia	10	312,547	312,543 (100.00%)	-	-	-	-	-	-	-	-	-	-	4 (0.00%)
13	2011, 2012	Philippines	11	88	87* (98.86%)	-	-	-	-	-	1 (1.14%)	-	-	-	-	-
14	2013	Philippines	8	5,332	6 (0.11%)	55* (1.03%)	8* (0.15%)	12* (0.23%)	-	-	5,251 (98.48%)	-	-	-	-	-

novo assembly of sequence reads into contigs, consensus sequences assembled from sequence reads by Trinity [18] followed by identification of the contigs in GENBANK by BLASTn and (ii) de novo assembly by meta-IDBA [19], followed by blast against complete genome

sequence data base. Novel pathogen classification followed the analysis procedure by BLASTx alignment against the non-redundant (nr) protein database as described previously [20] with the cut-off at 500 kb or greater in contig length with undetectable similarity to reference

**Table 1B**  
Viral pathogens identification of the pooled HEp-2 culture collected from South/South East Asia 2010–2013.

Pool No.	Year	Country	# Samples	#Viral Sequence Reads	Numbers of sequence reads identified as respiratory viruses (% identified sequence reads of identified virus/total viral sequence reads)											
					HSV-1	MuV	EV	CV	E	hRV	IFV A	IFV B	hPIVs	HAdV-1	HCoV	hMPV
15	2011, 2012	Nepal	11	672,015	5,444 (0.81%)	-	199,393 (29.67%)	294,728 (43.86%)	172,450 (25.66%)	-	-	-	-	-	-	-
16	2011	Bhutan	3	142,894	5,576 (3.90%)	-	-	137,318 (96.10%)	-	-	-	-	-	-	-	-
17	2013	Bhutan	6	2,581	2,572 (99.65%)	-	-	9* (0.35%)	-	-	-	-	-	-	-	-
18	2010, 2011	Thailand	17	302,989	4,085 (1.35%)	-	172,749 (57.01%)	56,448 (18.63%)	69,707 (23.01%)	-	-	-	-	-	-	-
19	2011	Thailand	19	436,239	8,088 (1.85%)	2* (0.00%)	78,443 (17.98%)	342,867 (78.60%)	6,839 (1.57%)	-	-	-	-	-	-	-
20	2011	Thailand	26	288,206	2,202 (0.76%)	226* (0.08%)	156,455 (54.29%)	124,593 (43.23%)	4,730 (1.64%)	-	-	-	-	-	-	-
21	2012	Thailand	3	31,554	12* (0.04%)	4* (0.01%)	31,538 (99.5%)	-	-	-	-	-	-	-	-	-
22	2012	Thailand	13	987,723	38,936 (3.94%)	2* (0.00%)	572,052 (57.92%)	259,961 (26.32%)	116,772 (11.82%)	-	-	-	-	-	-	-
23	2012	Thailand	12	438,254	51,063 (11.65%)	-	1,136 (0.26%)	372,335 (84.96%)	13,720 (3.13%)	-	-	-	-	-	-	-
24	2013	Thailand	13	7,951	7,951 (100.00%)	-	-	-	-	-	-	-	-	-	-	-
25	2013	Thailand	7	57,555	-	-	-	57,555 (100.00%)	-	-	-	-	-	-	-	-
26	2010-2013	Cambodia	11	443,407	69,740 (15.73%)	1,816* (0.41%)	202,754 (45.73%)	168,898 (38.09%)	199* (0.04%)	-	-	-	-	-	-	-
27	2011, 2012	Cambodia	12	15,564	15,564 (100.00%)	-	-	-	-	-	-	-	-	-	-	-
28	2012	Cambodia	11	81,818	6* (0.01%)	-	145* (0.18%)	60,676 (74.16%)	20,991 (25.66%)	-	-	-	-	-	-	-
29	2013	Cambodia	4	0	-	-	-	-	-	-	-	-	-	-	-	-
30	2011, 2012	Philippines	15	197,863	71,536 (36.15%)	-	192* (0.10%)	125,441 (63.40%)	292* (0.15%)	4 (0.00%)	-	-	-	-	398 (0.20%)	
31	2012	Philippines	8	931	10* (1.07%)	-	195* (20.95%)	445* (47.8%)	281* (30.18%)	-	-	-	-	-	-	-
32	2013	Philippines	2	0	-	-	-	-	-	-	-	-	-	-	-	-

A total of 105 hits with sequence reads identified viral pathogens from NCBI nt database (a total of 105 cells contain numbers) and 63 (63/105 (60%)) were confirmed positive with RT-PCR and rRT-PCR. Viral pathogens: herpes simplex virus type 1 (HSV-1), mumps virus (MuV), enterovirus (EV), coxsackievirus (CV), echovirus (E), Human rhinovirus (hRV), influenza A H3N2 (IFV A), influenza B (IFV B), human parainfluenza (hPIVs), human adenovirus (HAdV-1), human coronavirus (HCoV) and human metapneumovirus (hMPV). \*Less than 4.0% identified reads were observed as carried over among neighbor sequenced samples. Grey highlights indicate PCR positive confirmation the presence of the indicated pathogens. *Italic letters: No pathogens identified in the pools by PCR.*

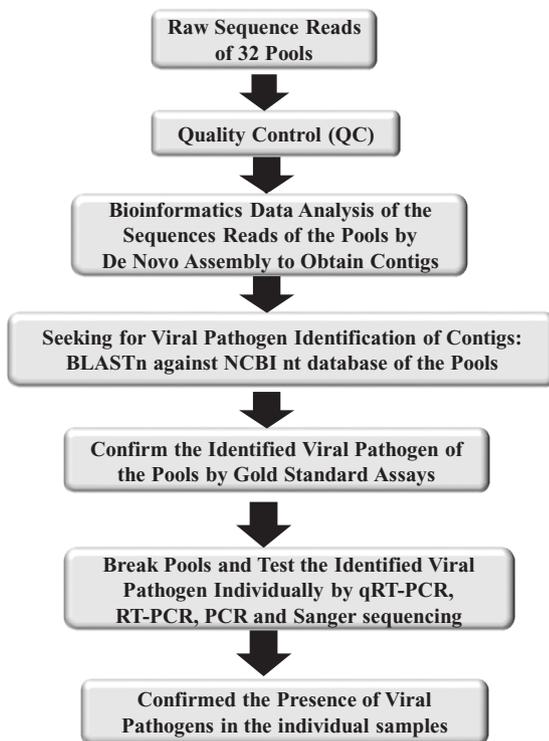


Fig. 2. Flow diagram illustrates steps for viral pathogen identification.

genome in NCBI nt database by BLASTn. Statistical analysis of the data comparisons was conducted with R Statistical Analysis Environment (version 2.15.3).

#### 4.6. Molecular assays for viral pathogen identification

Real-time RT-PCR, RT-PCR or PCR was used to confirm the WG-NGS results from CPE culture and clinical specimens. Table S1 displays the PCR and RT-PCR primers sequences used to confirm the WG-NGS results. The pan-EV RT-PCR [21] results were confirmed by Sanger sequencing at AITbiotech Pte Ltd. Company, Singapore. The accession numbers of our sequences in GENBANK were described in Table 3 with a small subset of samples previously published by Zhou et al., 2016. To determine the genetic variations and the relationships with the reference viruses, phylogenetic trees were constructed using Seaview (v4.0) and GTR+G+I parameter calculated by MEGA v. 6.0.

## 5. Results

### 5.1. WG-NGS on CPE positive cultures

During routine ILI surveillance in S/SE Asia in 2010–2013, approximately 4325 (33.6%) of 12,865 respiratory specimens collected were found to be positive for IFV by rRT-PCR and 8540 (66.4%) specimens were found to be negative (Fig. 1). Approximately 52.4% (4478 of 8540) of the IFV rRT-PCR negative specimens were cultured, 12.8% of which (572 of 4478) were found to induce CPE positive in either MDCK or HEp-2 cells. A total of 339 of 572 (59.3%) of these CPE positive cultures were negative by standard respiratory HAI and IFA assays, of which 324 (95.6%) (2–26 culture samples per pool) were pooled.

A total of 32 distinct pools were sequenced by WG-NGS to identify viral agents present in the supernatant following the flow diagram illustrated in Fig. 2. A total of 46.1 million sequence reads were generated from 3 sequencing runs. After selecting sequence reads with  $\geq 30Q$  scores and removing background sequences, 5.9 million sequence reads

remained for analyses. Sequence alignment using BLASTn in NCBI nt database identified viral sequences in 29 of 32 (90.6%) pools (Tables 1A and 1B). Pools from Cambodia 2013, Philippines 2012 and Philippines 2013 yield sequence reads identified host background and not viral pathogens by BLASTn in NCBI nt database.

Sequence reads aligning with HSV-1 were found to be most common (25 of 32 pools, 78%). In addition to HSV-1, 11 other viral sequences were detected in the pools, including mumps virus (MuV), enterovirus (EV), Coxsackievirus (CV), echovirus (E), human rhinovirus (hRV), IFV A, IFV B, hPIVs, human adenovirus (HAdV), human coronavirus (HCoV), and human metapneumovirus (hMPV). We verified the pools' results using conventional PCR and RT-PCR and were able to confirm the WG-NGS results in 63/105 cases (60%, Tables 1A and 1B, shaded cells). A discrepancy occurred in the Philippines 2012, where PCR/RT-PCR testing failed to corroborate the presence of any of the 4 viruses detected by WG-NGS. No pathogen identification was made by WG-NGS and PCR in three pools (Tables 1A and 1B, italic letters).

### 5.2. Confirmatory testings

The 310 culture supernatants that made up the 29 pools were further tested by pathogen-specific PCR, RT-PCR or rRT-PCR to validate the WG-NGS findings and to quantify the percentage of samples within each pool where viruses were present (Table 2). Two hundred and thirty-one of the 310 (74.5%) individual samples were positive using at least one of the pathogen-specific PCR tests verifying the pathogens identified by NGS in the pooled samples (Table 2). Approximately 43.3% (100/231) of the positive samples were respiratory pathogens. Of these, CVs were the most common respiratory viruses found, present in 31.0% (31/100) of all positive samples, followed by E, found in 22% samples. EV (16.0%), hRV, (3.0%), MuV (2.0%), hMPV (2.0%) and HCoV (1.0%) were all found at lesser frequencies (Table 2). Two samples positive for the pan-enterovirus (pan-EV) PCR could not be further typed and appear in Table 2 as pan-EV positive (2%). Also found were respiratory viruses that failed to be identified during the original screening for IFVA and HAdV due to few mismatched primer sequences (Fig. S2). The non-respiratory virus HSV-1 was the most common virus found in the individual supernatants, present in 172 of 310 samples (55.4%, Table S3). In at least 41 of these samples (23.8%) HSV-1 appeared as a co-infecting agent, most commonly co-present with enteroviruses. One sample collected from Philippines in 2012 contained HCoV OC43 strain co-infected with HSV-1. In addition, standard molecular procedures were conducted in selected samples and results validated the presence of the identified pathogens in both clinical specimens and CPE culture. Majority of clinical specimens were depleted and unavailable for this confirmation test.

### 5.3. Further characterization of enterovirus by Sanger sequencing

Seventy-one EV positive (22 from HSV-1 co-infected samples) samples from Nepal, Bhutan, Thailand, Cambodia and Philippines, were further subtyped within the EV genus utilizing the GTR+G+I model maximum likelihood tree (Fig. 3) using the Sanger sequences obtained from their PCR amplicons as previously described. The 71 positive pan-EV samples were found to be CV B5 (13/71; 18.3%), E 30 (13/71; 18.3%), CV B3 (10/71; 14.1%), E 7 (12/71; 16.9%), CV B4 (5/71; 7.0%), E 6 (5/71; 7.0%), hRV C (3/71; 4.2%), CV A21 (2/71; 2.8%), CV B1 (2/71; 2.8%), CV B2 (2/71; 2.8%), E 9 (2/71; 2.8%) and EV71 (2/71; 2.8%). Nine samples were sequenced individually at Viral Diseases Branch, Walter Reed Army Institute of Research (WRAIR) showing identical sequence and classifying the same enterovirus serotypes [22]. Five subtypes of the enterovirus genus and hRV species C were found in more than one country in SE Asia. The genetic distance among these samples of the same serotypes found in 5 countries were less than 0.2 (data not shown). One case of EV71 was detected from samples collected from Thailand in 2011.



**Table 3**  
GENBANK accession numbers of the enterovirus genus from our study.

No.	Organisms	Accession Numbers
1.	coxsackievirus A21 (CV A21)	KP062967- KP062968
2.	coxsackievirus B1 (CV B1)	KP062974- KP062975
3.	coxsackievirus B2 (CV B2)	KP062976- KP062977
4.	coxsackievirus B3 (CV B3)	KP062978- KP062987
5.	coxsackievirus B4 (CV B4)	KP062988- KP062992
6.	coxsackievirus B5 (CV B5)	KP062993- KP063005
7.	echovirus 6 (E 6)	KP063007- KP063011
8.	echovirus 7 (E 7)	KP063012- KP063023
9.	echovirus 9 (E 9)	KP063024- KP063036
10.	echovirus 30 (E 30)	KP0630006, KP063037
11.	enterovirus 71 (EV71)	KP062969- KP062970
12.	human rhinovirus C (hRV C)	KP062971- KP062973

\*The partial sequences these samples (KP062982, KP062997, KP063006, KP063009, KP063010, KP063011, KP063028, KP063030, KP062969) identically overlap with the full length sequences previously submitted by Zhou et al. [21] (KU574623, KU574625, KU574633, KU574628, KU574627, KU574626, KU574630, KU574629 and KU574619), respectively.

## 6. Discussions

A summary of the pathogens identified and the cases with still unclear etiology collected during routine respiratory surveillance 2010–2013 is illustrated in Fig. 4. Our analyses found HSV-1 to be the most frequent virus present in our supernatant, followed by EV genus and other common respiratory viruses, which was in agreement with the observation in Thailand reported by Zhou et al. [22]. The frequency of HSV-1 positive samples was most likely due to the exacerbation of existing HSV-1 infections caused by common respiratory infections in patients and the subsequent HSV-1 viral shedding in the upper respiratory track at the time of sample collection. Individuals with HSV-1 infections were routinely described as asymptomatic or subclinical [23]. It was estimated that HSV-1 seroprevalence was about 51% in SE Asia [24] and 33.3% in S Asia [25].

Viruses from the EV genus were the most common (71%) respiratory pathogens identified. Nine samples from Thailand 2010–2011 were sequenced individually and confirmed the present of the identical enterovirus serotypes sequences [22]. Despite the spread of EV-D68 globally [26], it was not detected in our surveillance study which enrolled adults 18–65 ages group during 2010–2013. Our study, with limited sample size detected a similar range of enterovirus serotypes (CV B2, CV B3, CV B4, E6 and E9) in SE Asian countries as other recent studies in Thailand [21,22]. These serotypes identified outside Thailand strains have genetic distance < 0.2 from the Thai serotypes. This suggests some temporal and spatial structure of both the Thai and, more broadly, the SE Asian EV populations, and thus may indicate some regional circulation of EV lineages and viral traffic between countries in SE Asia.

Overall, our study demonstrated that WG-NGS on pooled culture supernatants can be a useful tool to identify viral pathogens in clinical isolates. However, the identification of pathogens in the samples that yielded low sequence reads requires additional testing and controls. Our data (Tables 1A and 1B) showed total sequence reads without subtracting background sequence of the negative control experiment. The low sequence reads in several samples were likely generated from combination of cross talk among neighboring sequenced samples, contamination of barcodes and carries over from the previous runs. A specific example such as the Philippines 2012 pool contained sequence reads that were identical to Philippines 2011- 2012 pool, which was performed next to each other during library preparation. In addition, three pools detected only host background and not viral pathogens possibly due to low viral load, no viruses growing in the culture, or misidentifying the morphology of culture identified as CPE positive.

WG-NGS of pooled culture provided comprehensive information

from samples with unclear etiology after routine respiratory assays. Due to limitation of the study, a subset of the samples (2.9%) was selected for sequencing individually and the identity of the pathogens from sequence individually is in agreement of the results from the pooled samples. With the addition of WG-NGS, 29 of 32 (90.6%) pools were shown to contain viral pathogens identifiable using the GENBANK nt database. When broken down into individual samples, we were able to confirm the presence of other respiratory viruses in 100 of the 310 (32.3%) individual samples. WG-NGS significantly increased valid data information in addition to the standard respiratory assays. Furthermore, the WG-NGS and bioinformatics analysis were able to identify viral pathogens from pooled cultures containing up to 26 samples, which potentially makes this method time, labor and cost effective. This procedure allows screening of larger numbers of samples, with practical applications to surveillance or clinical studies.

However, the sensitivity of detection of this procedure may have some limitations [27]. Mainly pooled CPE positive cultures were tested as opposed to individually sequencing primary specimens. Without taking into consideration samples where HSV-1 was detected, approximately 67.8% of the samples in the pooled remained of unknown etiology after WG-NGS and bioinformatics data analysis (Fig. 4). A potential and likely significant limitation in this work was the cell lines selected for virus isolation. Our limitation to our two cell lines may have curtailed the type of viruses we were able to isolate and identify. Moreover, antibiotics required to maintain the MDCK and HEp-2 cultures may have also favored the growth of certain viruses over others.

Despite being a useful tool to screen for viral pathogens, the WG-NGS on pooled cultures may suffer of low resolution not only for pathogen identification but also for pathogen discovery. The WG-NGS utilizes random amplification which also amplifies host genome, increasing the complexity of the task of identifying pathogens generally present in low abundance in the clinical specimens and culture [28]. Long sequence reads with high depth of coverage (DOC) are optimal to identify diverse types of viruses from various types of clinical specimens [29–32]. For novel pathogen discovery, it would be required to apply advanced sample preparation with higher capacity host cell removal procedures directly from the primary clinical specimen and efficient pathogen discovery bioinformatics pipelines.

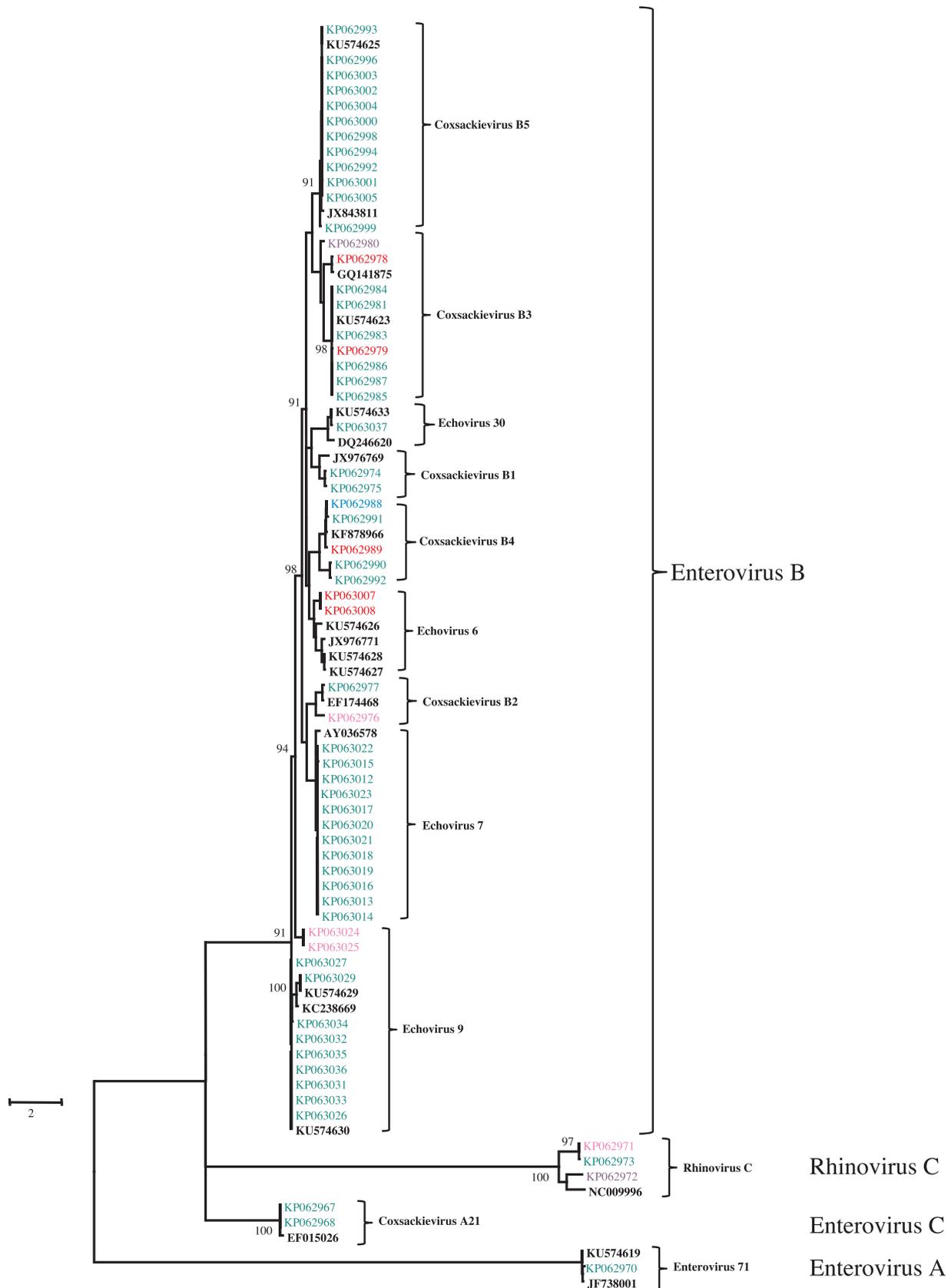
In summary, WG-NGS on pooled samples provides advantages as a screening tool to complement standard assays used during surveillance and diagnostic process. The WG-NGS does not require predefined target for identification of specific pathogens based on clinical presentation. In addition, the WG-NGS broadens the array, magnitude and complexity of pathogen detection since it is not limited by pathogen-specific primer and probe sequences. There are no requirements for the WG-NGS to update primer and probe sequences and the use of culture to isolate continuously. However, the technology provides complex results which require careful sample preparation, powerful sequencers, exhaustiveness of data analysis on well-planned bioinformatics pipelines and large databases for pathogen identification.

## Funding

Armed Forces Health Surveillance Center – Global Emerging Infections Surveillance and Response System (AFHSC-GEIS).

## Disclaimer

Material has been reviewed by the Walter Reed Army Institute of Research. There is no objection to its presentation and/or publication. The opinions or assertions contained herein are the private views of the author, and are not to be construed as official, or as reflecting true views of the Department of the Army, Department of the Navy or the Department of Defense.



**Fig. 3.** Maximum likelihood phylogenetic analysis of Enterovirus genus utilizing GTR + G + I model. Colors used in the phylogenetic tree represent: enterovirus references from GENBANK (black, n = 12); Nepal (pink, n = 4); Bhutan (blue, n = 1); Thailand (green, n = 59); Cambodia (red, n = 5) and Philippines (pink, n = 4). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

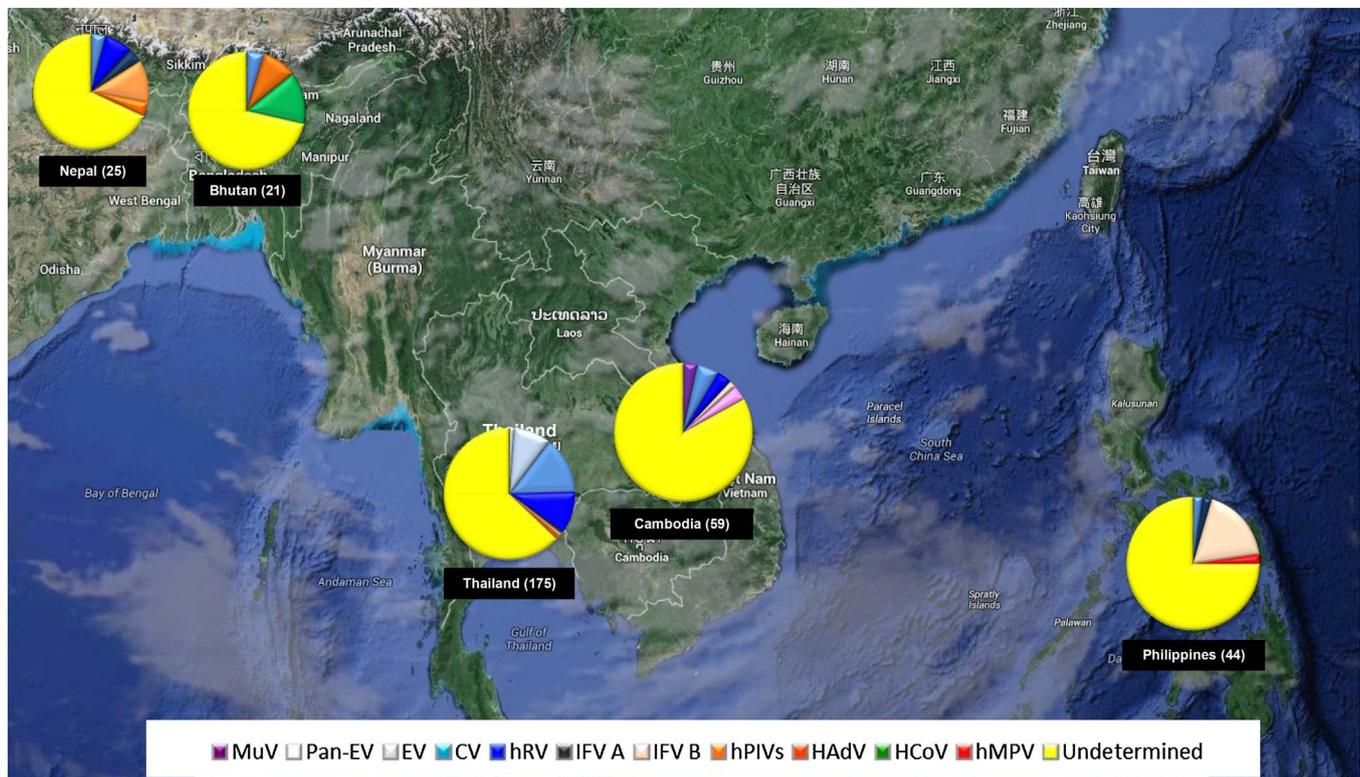


Fig. 4. Regional map displaying the distribution of respiratory viral pathogens identified by NGS from respiratory specimens collected during 2010–2013.

## Acknowledgements

We would like to thank Dr. Simon Pollett and Dr. Damon Ellison for comments on the manuscript; Dr. Kimberly Bishop-Lilly, Mr. Gregory Rice, Ms. Ragina Cer, Dr. Patrick Chain for advices in bioinformatics; Ms. Thidarar Intararit and the clinical research coordinator team for the support on the clinical information; Mr. Chitchai Hemachudha, Ms. Angkana Huang, Ms. Tipawan Thipwong and the project management section for specimen processing and demographic data on study subjects; Ms. Kamonthip Rungrojcharoenkit, Ms. Duangrat Mongkolsirichaikul and the virology and serology laboratory team for processing viral isolates, CPE and HAI. Mr. Kittinun Hussem from AFRIMS molecular section team and Ms. Phatcharin Chotchuang an undergraduate student from Burapha University for laboratory technical assistance on rRT-PCR, Tapestation and qPCR.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jcv.2017.07.004>.

## References

- [1] K.M. Peck, C.L. Burch, M.T. Heise, R.S. Baric, Coronavirus host range expansion and Middle East respiratory syndrome Coronavirus emergence: biochemical mechanisms and evolutionary perspectives, *Annu. Rev. Virol.* 2 (2015) 95–117.
- [2] H. Wang, Z. Zhang, Z. Chen, Y. Zhang, Q. Lv, X. An, et al., High genetic diversity and frequent genetic reassortment of avian influenza A(H9N2) viruses along the East Asian-Australian migratory flyway, *Infect. Genet. Evol.* 39 (2016) 325–329.
- [3] Q. Xiao, L. Ren, S. Zheng, L. Wang, X. Xie, Y. Deng, et al., Prevalence and molecular characterizations of enterovirus D68 among children with acute respiratory infection in China between 2012 and 2014, *Sci. Rep.* 5 (2015) 16639.
- [4] E.H. Graf, K.E. Simmon, K.D. Tardif, W. Hymas, S. Flygare, K. Eilbeck, et al., Unbiased detection of respiratory viruses by use of RNA sequencing-based metagenomics: a systematic comparison to a commercial PCR panel, *J. Clin. Microbiol.* 54 (2016) 1000–1007.
- [5] E.C. Holmes, The RNA virus quasispecies: fact or fiction, *J. Mol. Biol.* 400 (2010) 271–273.
- [6] C.W. Nelson, A.L. Hughes, Within-host nucleotide diversity of virus populations: insights from next-generation sequencing, *Infect. Genet. Evol.* 30 (2015) 1–7.
- [7] T. Takahashi, K. Maeda, T. Suzuki, A. Ishido, T. Shigeoka, T. Tominaga, et al., The first identification and retrospective study of Severe Fever with Thrombocytopenia Syndrome in Japan, *J. Infect. Dis.* 209 (2014) 816–827.
- [8] P.P. Afonso, P.M. Silva, L.C. Schnellrath, D.M. Jesus, J. Hu, Y. Yang, et al., Biological characterization and next-generation genome sequencing of the unclassified Cotia virus SPAn232 (Poxviridae), *J. Virol.* 86 (2012) 5039–5054.
- [9] T. Ho, R.R. Martin, I.E. Tzanetakos, Next-generation sequencing of elite beta germplasm and data analysis using a bioinformatics pipeline for virus detection and discovery, *Methods Mol. Biol.* 1302 (2015) 301–313.
- [10] T. Ho, I.E. Tzanetakos, Development of a virus detection and discovery pipeline using next generation sequencing, *Virology* 471–473 (2014) 54–60.
- [11] W. Rutvisuttinunt, W. Leelawiwat, P. Chinnawirotpisana, F. Mueanpaib, O. Kongpechsati, C. Klungthong, R. O'Connell, M. de Souza, I.K. Yoon, M. Curlin, S. Fernandez, Metagenomics Analysis of Plasma in HIV-infected Men Who Have Sex with Men in Bangkok, Thailand. HIV Research for Prevention 2014: AIDS Vaccine, Microbicide and ARV-based Prevention Science (HIV R4P) in Cape Town, South Africa 28–31 October. HIV Research for Prevention 2014: AIDS Vaccine, Microbicide and ARV-based Prevention Science (HIV R4P) in Cape Town, South Africa 28–31 October, 2014.
- [12] N. Fischer, D. Indenbirken, T. Meyer, M. Lutgehetmann, H. Lellek, M. Spohn, et al., Evaluation of unbiased next-generation sequencing of RNA (RNA-seq) as a diagnostic method in influenza virus-positive respiratory samples, *J. Clin. Microbiol.* 53 (2015) 2238–2250.
- [13] H.S. Ip, M.R. Wiley, R. Long, G. Palacios, V. Shearn-Bochsler, C.A. Whitehouse, Identification and characterization of Highlands J virus from a Mississippi sandhill crane using unbiased next-generation sequencing, *J. Virol. Methods* 206 (2014) 42–45.
- [14] E. Cabannes, C. Hebert, M. Eloit, Whole genome: next-generation sequencing as a virus safety test for biotechnological products, *PDA J. Pharm. Sci. Technol.* 68 (2014) 631–638.
- [15] C. Klungthong, P. Chinnawirotpisan, K. Hussem, T. Phonpakobsin, W. Manasatienkij, C. Ajariyakhajorn, et al., The impact of primer and probe-template mismatches on the sensitivity of pandemic influenza A/H1N1/2009 virus detection by real-time RT-PCR, *J. Clin. Virol.* 48 (2010) 91–95.
- [16] W. Rutvisuttinunt, P. Chinnawirotpisan, S. Simasathien, S.K. Shrestha, I.K. Yoon, C. Klungthong, et al., Simultaneous and complete genome sequencing of influenza A and B with high coverage by Illumina MiSeq Platform, *J. Virol. Methods* 193 (2013) 394–404.
- [17] W. Rutvisuttinunt, P. Chinnawirotpisan, B. Thaisomboonsuk, P. Rodpradit, C. Ajariyakhajorn, W. Manasatienkij, et al., Viral subpopulation diversity in influenza virus isolates compared to clinical specimens, *J. Clin. Virol.* 68 (2015) 16–23.
- [18] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, et al., Full-length transcriptome assembly from RNA-Seq data without a reference genome, *Nat. Biotechnol.* 29 (2011) 644–652.
- [19] Y. Peng, H.C. Leung, S.M. Yiu, F.Y. Chin, Meta-IDBA: a de Novo assembler for

- metagenomic data, *Bioinformatics* 27 (2011) i94–i101.
- [20] E.S. Lim, A. Reyes, M. Antonio, D. Saha, U.N. Ikumapayi, M. Adeyemi, et al., Discovery of STL polyomavirus, a polyomavirus of ancestral recombinant origin that encodes a unique T antigen by alternative splicing, *Virology* 436 (2013) 295–303.
- [21] J. Puenpa, J. Mauleekoonphairoj, P. Linsuwanon, K. Suwannakarn, T. Chieochansin, S. Korkong, et al., Prevalence and characterization of enterovirus infections among pediatric patients with hand foot mouth disease, herpangina and influenza like illness in Thailand, 2012, *PLoS One* 9 (2014) e98888.
- [22] Y. Zhou, S. Fernandez, I.K. Yoon, S. Simasathien, V. Watanaveeradej, Y. Yang, et al., Metagenomics study of viral pathogens in undiagnosed respiratory specimens and identification of human enteroviruses at a Thailand hospital, *Am. J. Trop. Med. Hyg.* 95 (2016) 663–669.
- [23] H.H. Handsfield, Public health strategies to prevent genital herpes: where do we stand? *Curr Infect Dis Rep.* 2 (2000) 25–30.
- [24] J.S. Smith, N.J. Robinson, Age-specific prevalence of infection with herpes simplex virus types 2 and 1: a global review, *J. Infect. Dis.* 186 (Suppl. 1) (2002) S3–28.
- [25] R. Kaur, N. Gupta, U.K. Baveja, Seroprevalence of HSV1 and HSV2 infections in family planning clinic attenders, *J. Commun. Dis.* 37 (2005) 307–309.
- [26] Centers for Disease Control and Prevention, Clusters of acute respiratory illness associated with human enterovirus 68—Asia, Europe, and United States, 2008–2010, *MMWR Morb. Mortal. Wkly. Rep.* 60 (2011) 1301–1304.
- [27] F. Thorburn, S. Bennett, S. Modha, D. Murdoch, R. Gunson, P.R. Murcia, The use of next generation sequencing in the diagnosis and typing of respiratory infections, *J. Clin. Virol.* 69 (2015) 96–100.
- [28] M. Lecuit, M. Eloit, The diagnosis of infectious diseases by whole genome next generation sequencing: a new era is opening, *Front. Cell. Infect. Microbiol.* 4 (2014) 25.
- [29] L. Barzon, E. Lavezzo, G. Costanzi, E. Franchin, S. Toppo, G. Palu, Next-generation sequencing technologies in diagnostic virology, *J. Clin. Virol.* 58 (2013) 346–350.
- [30] M.R. Capobianchi, E. Giombini, G. Rozera, Next-generation sequencing technology in clinical virology, *Clin. Microbiol. Infect.* 19 (2013) 15–22.
- [31] I. De Vlaminc, K.K. Khush, C. Strehl, B. Kohli, H. Luikart, N.F. Neff, et al., Temporal response of the human virome to immunosuppression and antiviral therapy, *Cell* 155 (2013) 1178–1187.
- [32] K.M. Wylie, K.A. Mihindukulasuriya, E. Sodergren, G.M. Weinstock, G.A. Storch, Sequence analysis of the human virome in febrile and afebrile children, *PLoS One* 7 (2012) e27735.