

RESEARCH ARTICLE

# A Next-Generation Sequencing Data Analysis Pipeline for Detecting Unknown Pathogens from Mixed Clinical Samples and Revealing Their Genetic Diversity

Yu-Nong Gong<sup>1,2</sup>, Guang-Wu Chen<sup>1,2,3</sup>, Shu-Li Yang<sup>1,4</sup>, Ching-Ju Lee<sup>1</sup>, Shin-Ru Shih<sup>1,3,4</sup>, Kuo-Chien Tsao<sup>1,3,4\*</sup>

**1** Department of Laboratory Medicine, Linkou Chang Gung Memorial Hospital, Taoyuan, Taiwan, **2** Department of Computer Science and Information Engineering, School of Electrical and Computer Engineering, College of Engineering, Chang Gung University, Taoyuan, Taiwan, **3** Research Center for Emerging Viral Infections, College of Medicine, Chang Gung University, Taoyuan, Taiwan, **4** Department of Medical Biotechnology and Laboratory Science, College of Medicine, Chang Gung University, Taoyuan, Taiwan

\* [kctsao@cgmh.org.tw](mailto:kctsao@cgmh.org.tw)



OPEN ACCESS

**Citation:** Gong Y-N, Chen G-W, Yang S-L, Lee C-J, Shih S-R, Tsao K-C (2016) A Next-Generation Sequencing Data Analysis Pipeline for Detecting Unknown Pathogens from Mixed Clinical Samples and Revealing Their Genetic Diversity. *PLoS ONE* 11 (3): e0151495. doi:10.1371/journal.pone.0151495

**Editor:** Zhi Wei, New Jersey Institute of Technology, UNITED STATES

**Received:** November 3, 2015

**Accepted:** February 29, 2016

**Published:** March 17, 2016

**Copyright:** © 2016 Gong et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** KCT received five grants from Chang Gung Memorial Hospital, Taoyuan, Taiwan (from No. CMRPG3C0731 to CMRPG3C0733, CLRPG3B0042 and CLRPG3B0043; <http://www.cgmh.org.tw>). GWC received one grant from the Ministry of Science and Technology, Taiwan (No. MOST-103-2221-E-182-022; <http://www.most.gov.tw/>). A total of 6 grants partially supported this work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Forty-two cytopathic effect (CPE)-positive isolates were collected from 2008 to 2012. All isolates could not be identified for known viral pathogens by routine diagnostic assays. They were pooled into 8 groups of 5–6 isolates to reduce the sequencing cost. Next-generation sequencing (NGS) was conducted for each group of mixed samples, and the proposed data analysis pipeline was used to identify viral pathogens in these mixed samples. Polymerase chain reaction (PCR) or enzyme-linked immunosorbent assay (ELISA) was individually conducted for each of these 42 isolates depending on the predicted viral types in each group. Two isolates remained unknown after these tests. Moreover, iteration mapping was implemented for each of these 2 isolates, and predicted human parechovirus (HPeV) in both. In summary, our NGS pipeline detected the following viruses among the 42 isolates: 29 human rhinoviruses (HRVs), 10 HPeVs, 1 human adenovirus (HAdV), 1 echovirus and 1 rotavirus. We then focused on the 10 identified Taiwanese HPeVs because of their reported clinical significance over HRVs. Their genomes were assembled and their genetic diversity was explored. One novel 6-bp deletion was found in one HPeV-1 virus. In terms of nucleotide heterogeneity, 64 genetic variants were detected from these HPeVs using the mapped NGS reads. Most importantly, a recombination event was found between our HPeV-3 and a known HPeV-4 strain in the database. Similar event was detected in the other HPeV-3 strains in the same clade of the phylogenetic tree. These findings demonstrated that the proposed NGS data analysis pipeline identified unknown viruses from the mixed clinical samples, revealed their genetic identity and variants, and characterized their genetic features in terms of viral evolution.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

Next-generation sequencing (NGS) technology has revolutionized virus discovery [1,2] and metagenomics [3,4] in the past decade. This technology provides a considerable number of reads from different pathogens without a priori knowledge about them. These considerable number of reads are assembled into contigs by computer algorithms [5,6]. The contigs are then searched for homologs in a database (usually BLASTN [7]). The top-hit organism/species fulfilling the search criteria (eg, sequence identity and e-value threshold) is used to “label” the contig, suggesting the existence of one particular organism/species. If assembled successfully, a contig can be long enough to reveal the entire viral genome of interest. However, the de novo assembled NGS reads often yield only short contigs that are partially aligned to or even do not match the database sequences [8], possibly because the tentative virus or similar strains have never been sequenced and thus do not appear in the database or because the virus mutated considerably (through point accumulation or recombination).

The Taiwan Virology Reference Laboratory Network was established by the Taiwan Centers for Disease Control in cooperation with 13 major medical centers for surveillance of viral infections [9]. Linkou Chang Gung Memorial Hospital (CGMH) plays a key role by processing a considerable number of viral specimens from Taiwan Taoyuan International Airport and collaborating hospitals and clinics in northern Taiwan for viral cultures. For example, 52 544 specimens were collected from 2000 to 2008 for investigating epidemiologic features and virus isolation of enteroviruses alone [10]. Furthermore, CGMH is responsible for detecting emerging and re-emerging viruses (eg, H1N1 [11] and H7N9 [12]) when confronting outbreaks. However, not all viruses can be detected using routine assays such as immunofluorescence assays (IFAs).

Human rhinovirus (HRV) and parechovirus (HPeV) are 2 viruses identified in this study. HRV, a member of the family of *Picornaviridae* and assigned as species of the genus *Enterovirus* in 2008, is a positive-sense, single-stranded RNA (ssRNA) virus with an approximately 7.2-kb genome. HRV causes common cold with mild symptoms. HPeV also belongs to the family of *Picornaviridae* but is assigned to a different genus, *Parechovirus*. It is also an ssRNA virus with an approximately 7.3-kb genome. HPeV primarily causes sepsis and central nervous system (CNS) diseases in infants, and still has other unproven clinical manifestations [13]. Laboratory diagnosis of HPeV is not routinely available in clinical practice and is still evolving; antiviral therapy is also being developed [14]. In Taiwan, among hundreds of viral isolates, isolates possibly containing HPeV might not be detected using standard enterovirus identification procedures [15]. Additional investigations are therefore required to clarify the characteristics of HPeVs for more effective clinical diagnosis. Current methods for detecting HPeV include real-time polymerase chain reaction (PCR) assays of stool, blood, respiratory swabs, and cerebrospinal fluid [16]. For HPeV typing, classification of the VP1 region is the current gold standard. However, because of an unknown viral load in clinical specimens, more sensitive VP1 primers for different HPeV genotypes are required [17]. Another issue is that primers are generally not available for all HPeV genotypes [18]. NGS technology involves producing genomic sequences from clinical samples without the need for designing specific primers, thus enabling the detection of unknown or untyped HPeVs.

This study detected viruses in 42 cytopathic effect (CPE)-positive cases isolated from 2008 to 2012 and revealed their genetic diversity by using NGS. Moreover, quality reads assembled into contigs were queried against known viral sequences in the database to predict the viruses present in the clinical sample. PCR-based detection or enzyme-linked immunosorbent assay (ELISA) was used to confirm the predicted viruses. Among the catalog of possible viruses obtained from these NGS experiments, we focused on identification of HPeVs because of their

reported clinical significance; furthermore, we explored their genomic context by mapping NGS reads onto them. Phylogenetic and recombination analyses were finally performed to discern the molecular evolution of the identified Taiwanese HPeVs.

## Materials and Methods

### Ethics statement

This study was approved by the Institutional Review Board (IRB) of Chang Gung Medical Foundation, Linkou Medical Center, Taoyuan, Taiwan (Approval No. 100-4378B).

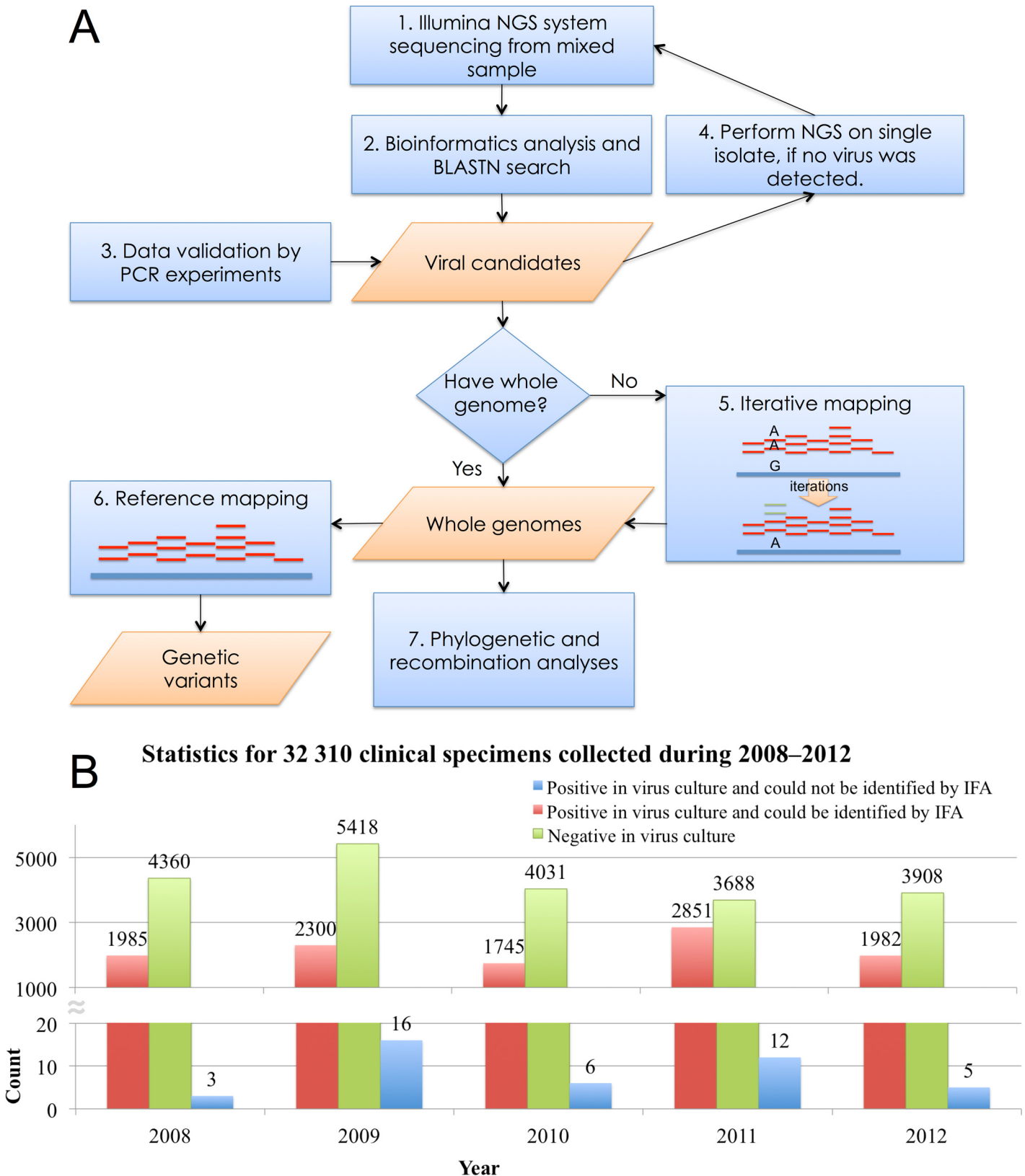
### Overall methodology for NGS pipeline

We implemented an NGS data analysis pipeline to identify unknown pathogens, assemble whole genomes, and investigate viral diversity from mixed/single clinical samples. [Fig 1A](#) shows the flowchart of the overall methodology. In brief, the overall methodology included the following steps: 1) NGS was performed using the Illumina system for collected clinical samples, 2) NGS reads were preprocessed using bioinformatics tools, and sequence homologs of assembled contigs were searched for BLASTN, 3) NGS diagnoses were validated using PCR or ELISA, 4) NGS analysis was performed on single isolate, if no virus was detected in PCR or ELISA confirmation, 5) an iterative mapping was implemented to obtain viral genome, if no genome was provided, 6) reference mapping was performed to identify genetic variants, and 7) genetic diversity was investigated using phylogenetic and recombination analyses. This platform is detailed in the following sections.

### Sample collection, virus isolation and RNA extraction

In total, 32 310 clinical specimens were received by CGMH for viral culturing from 2008 to 2012 ([Fig 1B](#)). Throat (TH) and nasopharyngeal (NP) swab specimens were inoculated into conventional tube cultures of MDCK, MK2, MRC-5 and RD cells. Rectal (RCT) swab specimens were inoculated into MK2, MRC-5, RD and A549 cells. These cell lines were obtained from ATCC (Manassas, Virginia, USA) and maintained in Minimum Essential Media (MEM) with 10% fetal bovine serum (FBS) and 1% antibiotic. The inoculation medium of MDCK and MK2 cells were MEM medium with 2 µg/ml trypsin and antibiotic, and of RD and MRC-5 cells were MEM medium with 2% FBS and antibiotic. Conventional culture tubes were incubated at 35°C with 5% CO<sub>2</sub>, and examined daily for CPE for two weeks. Infected cells were harvested and fixed with cold acetone for virus identification by IFA. Specific antibodies were used for detecting commonly seen viruses in clinical laboratory, as shown in [S1 Table](#). In addition, serotype-specific monoclonal antibodies were further employed for identifying enterovirus serotypes, if specimens were enterovirus positives.

Among these 32 310 clinical specimens in this study, more than 30% of the isolates were identified by IFA as positive cases (red bar in [Fig 1B](#)) every year. However, no viruses were detected by using IFA in 42 cultured samples (blue bar in [Fig 1B](#)), including 36 TH, 2 NP and 4 RCT swab specimens. These 42 isolates were centrifuged at the maximum speed (13 500 rpm) for 10 minutes in an eppendorf microcentrifuge tube for removal of residual cells and mitochondria and filtrated through a 0.22-µm filter for improving the sensitivity and reducing the background for subsequent analyses. Viral RNA extraction was performed in 140 µL of cell culture by using the QIAamp® Viral RNA Mini Kit (Qiagen, Valencia, USA). To extract nucleic acids, carrier RNA was replaced using linear acrylamide (Life Technologies, USA) as a precipitation reagent in the QIAamp® Viral RNA Mini Kit, according to previously published methods [[19](#)]. Extracted RNA eluted using 30 µL of elution buffer was stored at -70°C. Strain names



**Fig 1. A next-generation sequencing data analysis pipeline for clinical specimens.** A NGS data analysis pipeline was implemented to identify unknown pathogens, to reveal their genetic identity and variant, and to explore viral evolution from clinical samples. (A) Overall methodology. (B) Statistics for 32 310

clinical specimens collected during 2008–2012 in Linkou Chang Gung Memorial Hospital. The bar chart presents 42 cases (in blue) that could be cultured but not identified by IFA, 10 863 cases (in red) that could be cultured and identified by IFA and 21 405 cases (in green) that could not be cultured.

doi:10.1371/journal.pone.0151495.g001

(eg, TW-00158-2011) for these 42 isolates consist of 3 parts: country of sample collection (Taiwan, abbreviated as TW), a running number (eg, 00158), and the year of isolation (eg, 2011).

## Next-generation sequencing by using the Illumina system and bioinformatics analysis

To reduce the sequencing cost, the 42 specimens were pooled into 8 groups of 5 or 6 samples. The cDNA and library of mixed samples were synthesized using Ovation<sup>®</sup> RNA-Seq System V2 and Ovation<sup>®</sup> Ultralow Library Systems (NuGEN, San Carlos, CA, USA) according to previously published methods [20]; these were used by the Illumina system to generate raw data of 1.5 gigabytes for each mixed sample. Preprocessing of the NGS reads was performed using CLC Genomics Workbench (version 7.5) [21] and BLASTN (version 2.2.22). In brief, we first trimmed the raw data by removing low-quality and short reads, followed by merging all the paired reads into single reads. Second, paired reads that could not be merged and merged single reads were filtered by RNA-seq analysis to remove human genome fragments. Third, the remaining reads were assembled de novo to construct contig sequences. Finally, BLASTN was used to search the assembled contigs against the nucleotide sequence (NT) database of the National Center for Biotechnology Information (NCBI) downloaded from <ftp://ftp.ncbi.nih.gov/blast/db/>. Viral candidates were identified using the reported top BLASTN hits for each of the queried contig sequences fulfilling the search criterion of an e-value threshold of  $10^{-5}$ . A queried contig was required to have an average read depth of at least 20, which is the recommended lower bound for revealing a whole viral genome in metagenomic NGS data [22].

## PCR or ELISA of clinical isolates

Each of the 42 clinical isolates was tested using PCR or ELISA to confirm the findings of the NGS data analysis. For PCR, viral RNA was amplified and detected using virus-specific primers or probes. Amplified products, if necessary, were sequenced using the BigDye Terminator Cycle Sequencing Kit.

Subtype-specific primers targeting the VP1 region were designed to detect HPeV-1 (primer set F1382-R3140 in Table 1). For HPeV-3 and -4, a universal primer set F1382-R3072 was designed, and the obtained VP1 sequence was queried against NT database of NCBI to determine the subtype. For HAdV detection and typing, type-specific primers and probes [23–25] were used to detect the amplified PCR product without sequencing. The detection of HRVs and echoviruses was directly performed using EV CODEHOP primers [26] to obtain the EV sequence, followed by querying the database for the serotype. For rotavirus detection, specimens were tested for the presence of the viral antigen by using the RIDASCREEN Rotavirus ELISA Kit according to manufacturer instructions.

## Amplifying and Sanger sequencing of the HPeV genomes

To explore the genetic variants among the identified HPeVs, we designed primers to obtain their complete coding sequences, as shown in Table 1. Viral RNA was amplified and sequenced using these primers. Eight HPeV genomes obtained in this study have been deposited in GenBank database with accession numbers KT626005 to KT626012.

**Table 1. Primer designs for the HPeV genome.**

Primer sets	Forward (F) and reverse (R) primers with start positions	Sequence (from 5' to 3')
1	F648	ACGTCTAGTGGGCCAAACC
	R2232	GAACCCAYGAATTTTGG
2	F1382	CCCTATGTTGCWGACACAAACTA
	<sup>a</sup> R3140	CTGTYTTGAAAATGTCATCTGA
3	F1382	CCCTATGTTGCWGACACAAACTA
	<sup>b</sup> R3072	CATAGTGCTTTRARAAACCCCT
4	<sup>a</sup> F3022	AGAYCAGAGYCCATATG
	R4691	ACTTCCCAGCAYTCACCAGT
5	<sup>b</sup> F3074	GGGTTTTTAYAARCACTAT
	R4691	ACTTCCCAGCAYTCACCAGT
6	F4475	TTTATWGTWCCAATGGCACACA
	R5933	GCTGGGCCCATYTTTRAC
7	F4356	CWGCTAGTGAGTTYATGGAT
	R6428	TCATARATTTCCATCATGATCAT
8	F5768	ATGCAYATAGCTGGRAATGG
	R7321	TTTGGTATGTCCAATATTCCAA

This table lists the 8 primer sets used for Sanger sequencing of the HPeV genome. Their start positions were based on reference strains (accession number S45208). Superscripts a and b are marked for specific primers of HPeV-1 and HPeV-3 and -4 based on the VP1 region, respectively. Other primers without superscripts are universal primers for all three types.

<sup>a</sup>Specific primers for HPeV-1

<sup>b</sup>Specific primers for HPeV-3 and -4

doi:10.1371/journal.pone.0151495.t001

## Reference mapping for genetic variant identification

To investigate the read depth and detect genetic variants of candidate viral genomes, we performed reference mapping using genomic templates with bowtie2 (version 2.2.5) [27] with default settings. Position-specific read counts and nucleotide compositions obtained after the read mapping on viral genome were examined, and a genetic variant was located where the primary nucleotide alphabet dominated less than 75% [28] and the read depth was at least 20.

## Iterative mapping for revealing viral genomes

Whole genomes by Sanger sequencing for the viruses of interest may fail due to low viral load, temperature, and so on. Therefore, no genomic template can be used for NGS read mapping. An iterative mapping procedure was implemented to construct viral genome, as follows.

Step 1. Select quality viral contigs by a depth of at least 20, and their top-hit BLASTN subjects

Step 2. Generate an initial template for reference mapping

- a. If quality contigs cover whole genome, a consensus sequence of these contigs will be the initial template.
- b. If quality contigs only cover partial genome, a consensus sequence of their BLASTN subjects was used for filling up missing regions. This modified sequence will be the initial template.

- Step 3. Do reference mapping
- Step 4. Nucleotide compositions of mapped reads at each genomic position were examined and used to substitute the nucleotide on the template sequence if a mismatch occurred. The modified template was then used as a new reference for the next round of reference mapping in Step 3 until no more substitutions were made to the template.
- Step 5. Output the terminal template as target genome.

## Phylogenetic and recombination analyses

The viral sequences identified in this study, together with reference genomes retrieved from NCBI, were aligned using Clustal Omega (version 1.2) [29] for subsequent phylogenetic and recombination analyses. Phylogenetic analysis was performed using MEGA6 [30], in which an evolutionary tree was inferred using the maximum likelihood method based on the Jukes–Cantor model [31]. Initial trees for the heuristic search were obtained by applying the neighbor-joining method to a matrix of pairwise distances estimated using the maximum composite likelihood (MCL) approach. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. The bootstrap value was set to 1000. All positions containing gaps were eliminated. Recombination was detected using SimPlot (version 3.5.1) [32] with a window and step size of 600 and 20 nucleotides, respectively.

## Results

### Taxonomic classification and detection of potential viral types by using NGS

The 42 clinical specimens were pooled into 8 groups (NGS 1 to 8; [Table 2](#)). NGS was conducted for each group of mixed samples. Contigs assembled from each of these NGS experiments were cataloged into viral types by taxonomic classification based on BLASTN. In total, 361 quality viral contigs (with a depth of at least 20) from all 8 NGS experiments were queried, which predicted 47 potential viral types and their BLASTN reports as shown in [Table 2](#) and [S1 File](#), respectively. HRVs were detected in all these NGS data sets, except for NGS 2. In particular, NGS 4, 6, 7, and 8 only contained various subtypes of HRV type A. HPeVs of different types were observed in NGS 1, 2, 3, and 5. Other predicted viruses included echovirus type 3 (NGS 1), human adenovirus (HAdV) type 3 (NGS 1 and 2), and rotavirus (NGS 3 and 5).

The 361 viral contigs assembled from NGS reads ranged from 62 bp to 10 290 bp. Although few long contigs might completely or nearly span the full genome, many short ones mapped only partially to the targeted viruses. We calculated the genome coverage (%) of the mapped contigs; the coverage is shown in parentheses in [Table 2](#). Some viruses had a coverage of 100%, such as predicted HPeV-1 and echovirus 3 in NGS 1. Only a small proportion had a low coverage. For example, HAdV-3 in NGS 1 showed only a genome coverage of 1%. In brief, 35 of the 47 identified viral types had a genome coverage of 64% or higher; these are shown in boldface font in [Table 2](#).

### PCR or ELISA of each of 42 clinical isolates

Following the analysis of the NGS reads for the mixed samples, virus-specific molecular diagnoses for each of the 42 isolates were performed depending on the NGS-predicted viruses included in [Table 2](#) for validating their presence. Viruses tested included HRV, HPeV, HAdV, rotavirus, and echovirus (a total of 47 viruses predicted from the 8 NGS experiments). All isolates tested positive for only one virus, except for TW-01662-2010, which tested negative for

**Table 2. Preliminary NGS prediction and PCR or ELISA detection.**

NGS group: Predicted viruses with genome coverage (%)	Strain names of isolates in mixed sample	PCR/ELISA detections by isolates					
		HAdV-2/3/7	HPeV-1/3/4	Rota	HRV/Echo	Final result	
NGS 1: <b>HPeV-1 (100), Echo 3 (100), HRV-B52 (99), HRV-A21 (97), HRV-A81 (97),</b> HAdV-3 (1)	TW-00158-2011	–	–	–	<b>HRV-A21</b>	<b>HRV-A21</b>	
	TW-70307-2011	–	–	–	<b>Echo-3</b>	<b>Echo-3</b>	
	TW-90330-2011	–	–	–	<b>HRV-B52</b>	<b>HRV-B52</b>	
	TW-50192-2012	–	<b>HPeV-1</b>	–	–	<b>HPeV-1</b>	
	TW-96046-2012	–	–	–	<b>HRV-A81</b>	<b>HRV-A81</b>	
	TW-00032-2011	–	<b>HPeV-4</b>	–	–	<b>HPeV-4</b>	
	TW-03067-2011	–	<b>HPeV-3</b>	–	–	<b>HPeV-3</b>	
	TW-71157-2011	–	<b>HPeV-1</b>	–	–	<b>HPeV-1</b>	
NGS 2: <b>HPeV-1 (95), HAdV-3 (93), HPeV-4 (82), HPeV-3 (81)</b>	TW-71170-2011	<b>HAdV-3</b>	–	–	–	<b>HAdV-3</b>	
	TW-01679-2012	–	<b>HPeV-1</b>	–	–	<b>HPeV-1</b>	
	NGS 3: <b>HPeV-1 (99), HRV-A56 (98), HRV-A21 (97), HRV-B4 (64), HRV-A20 (44),</b> HRV-A12 (35), Rotavirus (13)	TW-71236-2008	–	–	–	<b>HRV-A56</b>	<b>HRV-A56</b>
		TW-01319-2010	–	<b>HPeV-1</b>	–	–	<b>HPeV-1</b>
		TW-71397-2010	–	–	–	<b>HRV-B04</b>	<b>HRV-B04</b>
		TW-71422-2010	–	–	–	<b>HRV-A21</b>	<b>HRV-A21</b>
	TW-71594-2010	–	<b>HPeV-1</b>	–	–	<b>HPeV-1</b>	
	NGS 4: <b>HRV-A20 (97), HRV-A12 (89), HRV-A49 (89)</b>	TW-01662-2010	–	–	–	–	–
TW-00663-2011		–	–	–	<b>HRV-A20</b>	<b>HRV-A20</b>	
TW-00824-2011		–	–	–	<b>HRV-A20</b>	<b>HRV-A20</b>	
TW-70142-2011		–	–	–	<b>HRV-A12</b>	<b>HRV-A12</b>	
TW-71368-2011		–	–	–	<b>HRV-A49</b>	<b>HRV-A49</b>	
NGS 5: <b>HRV-A63 (97), HRV-A21 (96), HRV-A28 (87), HPeV-1 (87), Rotavirus (88),</b> HPeV-4 (58), HRV-A49 (45), HPeV-3 (19), HRV-A20 (9), HRV-A12 (2)	TW-02603-2008	–	–	<b>Rota</b>	–	<b>Rota</b>	
	TW-02680-2008	–	<b>HPeV-1</b>	–	–	<b>HPeV-1</b>	
	TW-96061-2010	–	–	–	<b>HRV-A21</b>	<b>HRV-A21</b>	
	TW-02547-2011	–	–	–	–	–	
	TW-50289-2012	–	–	–	<b>HRV-A63</b>	<b>HRV-A63</b>	
	TW-50733-2012	–	–	–	<b>HRV-A28</b>	<b>HRV-A28</b>	
	NGS 6: <b>HRV-A94 (98), HRV-A11 (97), HRV-A23 (96), HRV-A12 (89)</b>	TW-00421-2009	–	–	–	<b>HRV-A94</b>	<b>HRV-A94</b>
		TW-02960-2009	–	–	–	<b>HRV-A11</b>	<b>HRV-A11</b>
TW-03022-2009		–	–	–	<b>HRV-A12</b>	<b>HRV-A12</b>	
TW-03511-2009		–	–	–	<b>HRV-A12</b>	<b>HRV-A12</b>	
TW-70387-2009		–	–	–	<b>HRV-A23</b>	<b>HRV-A23</b>	
NGS 7: <b>HRV-A55 (97), HRV-A95 (97), HRV-A43 (94), HRV-A41 (88), HRV-A60 (86),</b> HRV-A11 (17)	TW-70426-2009	–	–	–	<b>HRV-A95</b>	<b>HRV-A95</b>	
	TW-70443-2009	–	–	–	<b>HRV-A43</b>	<b>HRV-A43</b>	
	TW-70656-2009	–	–	–	<b>HRV-A60</b>	<b>HRV-A60</b>	
	TW-71633-2009	–	–	–	<b>HRV-A55</b>	<b>HRV-A55</b>	
	TW-71682-2009	–	–	–	<b>HRV-A41</b>	<b>HRV-A41</b>	
NGS 8: <b>HRV-A7 (97), HRV-A55 (97), HRV-A16 (94), HRV-A20 (96), HRV-A41 (91),</b> HRV-A8 (58), HRV-A12 (21)	TW-71693-2009	–	–	–	<b>HRV-A16</b>	<b>HRV-A16</b>	
	TW-71980-2009	–	–	–	<b>HRV-A20</b>	<b>HRV-A20</b>	
	TW-72087-2009	–	–	–	<b>HRV-A41</b>	<b>HRV-A41</b>	
	TW-72141-2009	–	–	–	<b>HRV-A55</b>	<b>HRV-A55</b>	
	TW-72216-2009	–	–	–	<b>HRV-A7</b>	<b>HRV-A7</b>	
	TW-72323-2009	–	–	–	<b>HRV-A20</b>	<b>HRV-A20</b>	

A preliminary NGS prediction of potential pathogens indicates the presence of 47 viral types based on quality contigs. The coverage of 35 viral types is higher than 65% (shown in boldface font and confirmed through PCR), and that of the other 12 viral types is 58% or less. The coverage was obtained by using the covered region of contig(s) divided by the genomic length. The genomic length of HAdV was 36 000 bp, and the lengths of HRV, HPeV, and Echo were 7250 bp. The total lengths of all 11 rotavirus segments were 18 555 bp. All regions covered by one or more contigs in each segment were summed and divided by the genomic length for calculating the coverage. Depending on our NGS-predicted viruses, PCR or ELISA was performed for validation. HAdV and HPeV were confirmed using specific PCR primers and following HRV and Echo by EV CODEHOP PCR primers and rotavirus by ELISA. The final results show 40 identified viruses: 29 HRV, 8 HPeV, 1 echovirus, 1 HAdV, and 1 rotavirus.

doi:10.1371/journal.pone.0151495.t002



EV CODEHOP in NGS 4 where 3 HRVs were predicted, and TW-02547-2011, which tested negative for HPeV, rotavirus, and EV CODEHOP in NGS 5 where 6 HRVs, 3 HPeVs, and 1 rotavirus were predicted. The final results of molecular diagnosis for each of the 42 isolates are listed in the right-most column of [Table 2](#). The presence of NGS-predicted viruses in [Table 2](#) with a genome coverage of 64% and higher in the isolates was confirmed using PCR or ELISA. All predicted viruses with a genome coverage of 58% or lower tested negative by PCR or ELISA.

## NGS experiments and viral detection for 2 isolates with unknown viruses

NGS was separately implemented for the 2 isolates in [Table 2](#) that tested negatives after all tests. For TW-01662-2010 (designated as NGS 9), 2 quality contigs (each of 2633 and 3730 bp and average depths of 35 and 49) matched positions 830 to 3462 of the HPeV-1 K129-93 genome (accession number GQ183022) and 3492 to 7077 of the HPeV-1 CAU10-NN genome (JX575746) on BLASTN. In other words, these 2 contigs encompassed most of the first half and nearly the complete second half of the HPeV-1 genome, suggesting that TW-01662-2010 is an HPeV-1 strain. Subsequently, we implemented the iterative mapping approach proposed in this study to alter the genomic content of this HPeV-1 template to obtain as many NGS reads as possible. In brief, a consensus sequence was generated from the 2 HPeV-1 genomes, which were BLASTN subjects obtained by querying the 2 contigs. This consensus sequence was used for assembling the missing regions to generate an initial template, because contigs covered nearly HPeV genome. Subsequently, the iteration process was started from this combined genome. After 5 iterations, a terminal genome was obtained with an average mapping depth of 50.1, which increased from 35 and 49 depths of two partial sequences in 2633 and 3730 bp, respectively. A total of 56 nucleotides were present in the ORF region corrected from the initial template to the terminal template sequence. Similarly, the NGS experiment (designated as NGS 10) of TW-02547-2011 resulted in one quality viral contig of 7249 bp and depth 90, which matched the HPeV-1 PicoBank strain (FM242866) in the database. An iterative procedure started with this contig (already covering the whole genome), and after 56 iterations, it ended with a contig of depth 96.7 based on the ORF region of the terminal template. No difference was observed in nucleotides between the 2 templates based on the ORF region. That is, all iterations were the cost for substituting nucleotides in 2 UTR regions. In addition, BLAST reports of 3 quality contigs from NGS 9 and 10 were summarized in [S1 File](#).

The iterative mapping approach in this study was implemented to alter the genomic content for obtaining viral genome by attracting as many NGS reads as possible. The depths and genomic content of the terminal templates in NGS 9 and 10 seemed to be moderately corrected from their respective initial contigs through this iterative mapping process. Because the final depths in NGS 9 and 10 were 50.1 and 96.7, respectively, which were greater than the generally accepted depth threshold of 20 in exploring virus metagenomics [22], we considered that these 2 HPeV-1 viruses were present in the 2 isolates; their CDS sequences are provided in [S2 File](#).

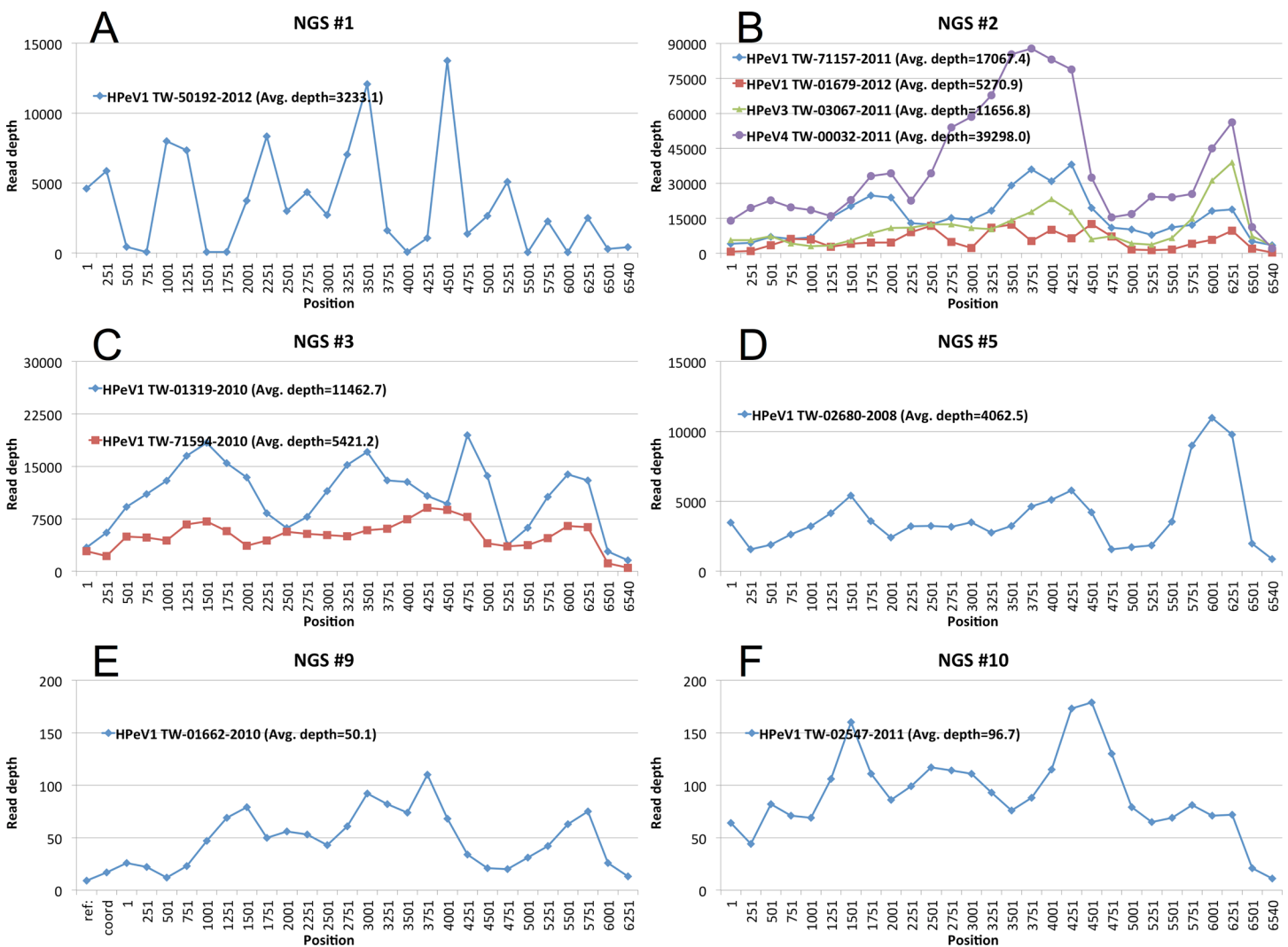
## Amplifying and Sanger sequencing of HPeV genomes from the isolates

Viruses detected among the 42 isolates included 27 HRV-A, 2 HRV-B, 8 HPeV-1, and 1 each of HPeV-3, HPeV-4, HAdV-3, rotavirus, and echovirus-3. Although HRVs were mostly detected in these clinical samples, we focused on the detected HPeVs and explored their genetic characteristics because of their clinical significance. The viral genome of each of the 8 PCR-confirmed HPeVs listed in [Table 2](#) was determined using Sanger sequencing, for which 8 primer sets as shown in [Table 1](#) were designed using the prototype strain Harris (accession number S45208). Eight genomes were obtained with lengths ranging from 6993 bp for TW-

00032-2011 to 7231 bp for TW-01679-2012, covering the entire ORF sequence and the partial UTR sequence. In particular, none of these HPeV primers were applicable to the 2 HPeV-1 viruses predicted from NGS 9 and 10. These 2 HPeV-1 genomes were revealed by NGS assembly analysis. In summary, the genetic characteristics of 10 HPeV genomes, including 8 determined by Sanger sequencing and 2 assembled from NGS analysis, were investigated.

### Read depth and genetic variants of HPeV genomes

NGS reads from each of the 10 experiments were respectively mapped to each of the 10 HPeV genomes, including 8 obtained by PCR and the 2 from analyzing NGS reads alone. This was performed to assess the overall depth distribution along the genome and to identify genetic variants exhibiting heterogeneous nucleotide compositions. Fig 2A–2F shows the read coverage distribution along the ORF of the reported HPeV genomes for NGS 1 (one HPeV-1 strain), 2 (two HPeV-1 strains, and one strain each of HPeV-3 and HPeV-4), 3 (two HPeV-1 strains), 5

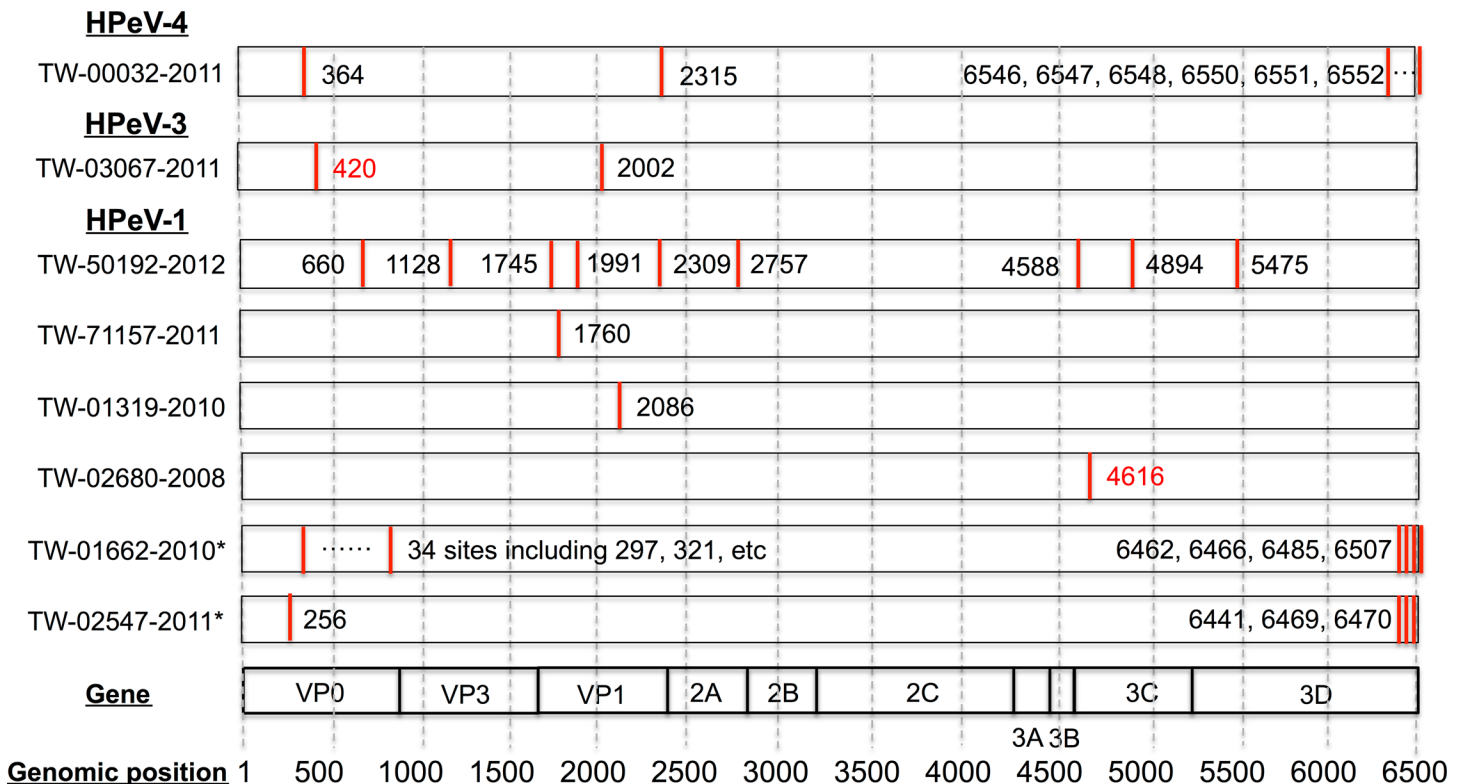


**Fig 2. Read distributions of HPeV genomes revealed in NGS experiments.** (A)–(D) Read distributions for 8 genomes obtained using PCR. NGS data sets 1, 2, 3, and 5 were each mapped to 1, 4, 2, and 1 HPeV genomes, respectively, generated using PCR. (E) and (F) Read distributions for 2 genomes revealed by NGS contig assembly, BLASTN search, and iterative mapping (as described in the Materials and Methods section).

doi:10.1371/journal.pone.0151495.g002

(one HPeV-1 strain), 9 (one HPeV-1 strain), and 10 (one HPeV-1 strain), respectively. The depth distributions in Fig 2A–2D showed non-zero coverage across the entire ORF, and an average depth of at least 3000 was obtained. The 2 genomes obtained from assembling NGS reads and iterative mapping also showed non-zero distributed depth coverage in the ORF (Fig 2E and 2F), although their depths were much more shallow than those were observed in the other 8 HPeV genomes.

We examined the mapped reads on each of the 10 genomes in Fig 2 and revealed their genetic variants. As mentioned earlier, a genetic variant was defined as a nucleotide position where the primary nucleotide alphabet appeared in less than 75% of the mapped reads with a read depth of at least 20. Fig 3 shows the genomic locations based on CDS that were graphed by the HPeV type. The HPeV-4 genome (TW-00032-2011) contained 8 variants: 1 in VP0, 1 in 2A, and 6 others at the 3' end of 3D. The HPeV-3 genome (TW-03067-2011) displayed 2 variants: 1 in VP0 and 1 in VP1. The one at position 420 of VP0 (marked in red in Fig 3) consisted of A (74.6%) and C (25.1%) and is a nonsynonymous substitution leading to an amino acid change from Glu to Asp. Six of the eight TW HPeV-1 genomes collectively displayed 54 genetic variants, among which 42 were abundant in VP0 and 3D in 2 in-silico assembled genomes (TW-01662-2010 and TW-02547-2011 in Fig 3) from NGS. The other 12 genetic variants were distributed on 4 TW HPeV-1 genomes obtained Sanger sequencing. The other nonsynonymous substitution at the nucleotide position 4616 in 3C of TW-02680-2008 consisted of C (69.1%) and T (30.4%) and led to an amino acid change from Pro to Leu. S3 File provides comprehensive information of these genetic variants, including their Sanger sequencing bases, NGS



**Fig 3. Genetic variants with heterogeneous genetic populations.** In total, 64 genetic variants were detected from 6 HPeV-1 (asterisks indicate 2 predicted genomes), 1 HPeV-3, and 1 HPeV-4 genomes, including 38 in VP0, 14 in 3D, 7 in VP1, 3 in 3C, 1 in 2A, and 1 in VP3. Among 64 variants, 8 and 2 were detected from HPeV-4 and HPeV-3, respectively, and the others from HPeV-1. Two nonsynonymous mutations at positions 4616 in HPeV-1 (TW-02680-2008) and 420 in HPeV-3 (TW-03067-2011) are marked with red text.

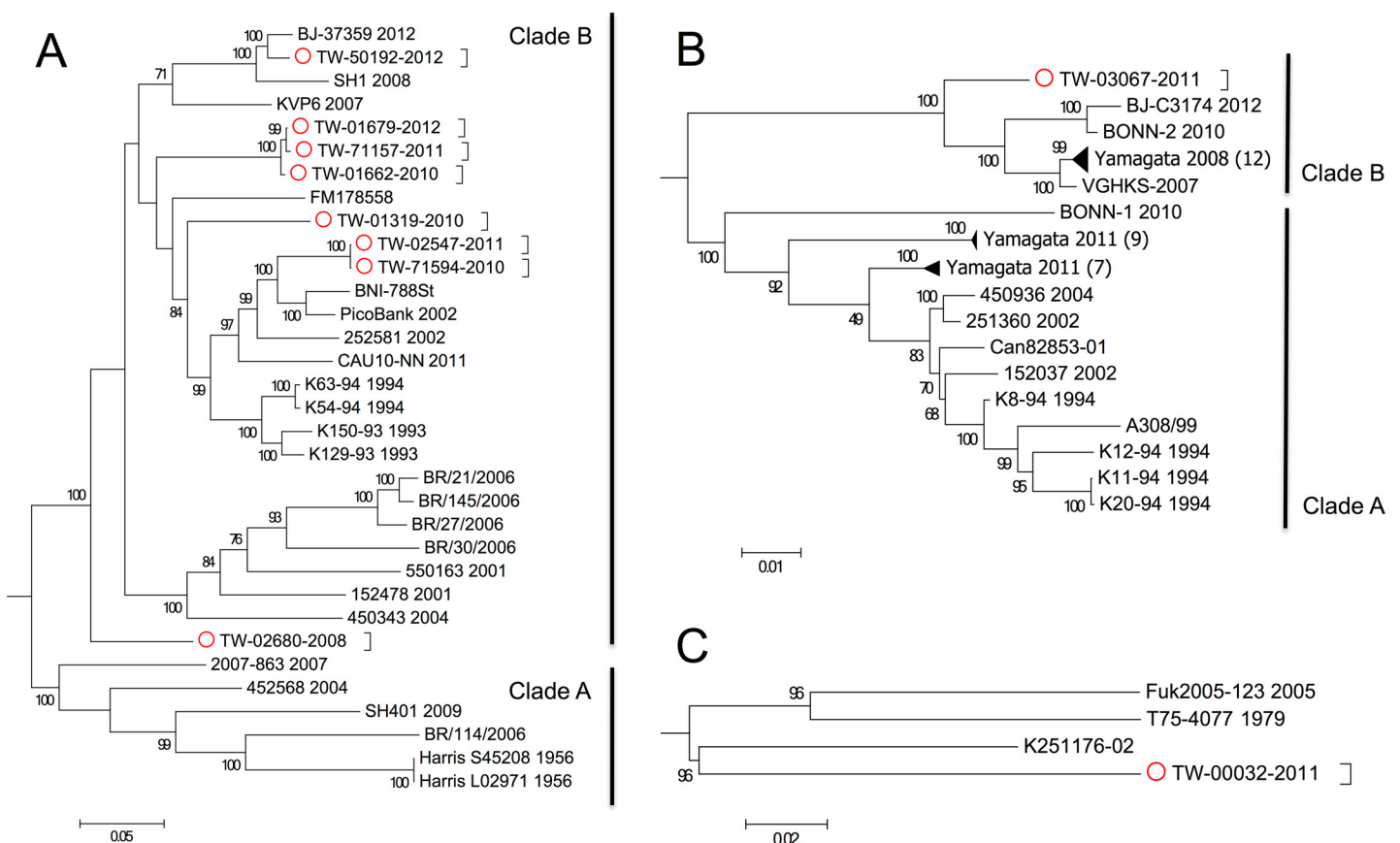
doi:10.1371/journal.pone.0151495.g003

read depths and compositions, NGS read consensus in nucleotides, and the translated amino acids from the corresponding codons containing these genetic variants.

### Genetic evolution of HPeV genomes

All HPeV genomes from NCBI were retrieved in March 2015 for assessing the sequence diversity and evolution of the 10 TW HPeV strains obtained in this study. The multiple sequence alignment revealed one distinct 6-bp deletion at positions 97 to 102 (in VP0 based on HPeV-1 CDS of 6540-nt long) in the HPeV-1 strain TW-02680-2008. This deletion resulted in a missing 2-aa short peptide “GN” in the translated protein sequences, as shown in [S1 Fig](#).

[Fig 4A](#) shows the phylogenetic tree based on 33 complete HPeV-1 ORF sequences, including the 8 TW sequences identified in this study. The most distinct pairwise percent identity among all 33 HPeV-1 sequences (or 32 sequences excluding the Harris strain isolated in 1956) was 79%, demonstrating genetic heterogeneity within this particular virus within a span from 1956 (or 1993) to 2012. All TW HPeV-1 genomes were classified into Clade B, within which the pairwise percent identities were 86% or higher, suggesting that TW strains were similar to the already published Clade-B viruses in the database. One of the TW strains, TW-02680-2008, showed an average identity of 86.3%. Similarly, comparing the other TW strains to the Clade-B



**Fig 4. Phylogenetic analysis of genomes of HPeV types 1, 3, and 4.** Genomes of all HPeV types 1, 3, and 4 from NCBI were retrieved in March 2015 for phylogenetic analysis with the 10 genomes (red circles) obtained in this study. (A) 33 HPeV-1 ORF sequences, including 8 from this study, were used to generate a tree branching into Clades A and B. (B) 42 HPeV-3 ORF sequences, including one from this study (TW-03067-2011), were used to generate a tree branching into Clades A and B. The 3 groups of Yamagata strains were compressed by triangles; their sequence counts are shown in parentheses. (C) 4 HPeV-4 ORF sequences, including one from this study (TW-00032-2011), were used to generate a tree.

doi:10.1371/journal.pone.0151495.g004

strains, 7 average identities were from 87.3% to 88% and were slightly stronger than that of TW-02680-2008 (86.3%). TW-02680-2008 was distantly located from the other 26 Clade-B viruses in Fig 4A. A phylogenetic tree of TW-03067-2011 with other 41 HPeV-3 viruses retrieved from database is graphed in Fig 4B. The most distinct pairwise percent identity among all these HPeV-3 sequences was 88%. This demonstrated that HPeV-3 viruses exhibited less selection pressure in a similar span from 1994 to 2012 in comparison with HPeV-1 population, which displayed a much smaller percent identity of 79%. TW-03067-2011 had least 95% identity to the other HPeV-3 viruses in Clade B in Fig 4B. Fig 4C shows the tree for an HPeV-4 strain (TW-00032-2011) identified in this study with 3 HPeV-4 reference genomes available from NCBI. TW-00032-2011 showed 83.0%–83.6% identity compared with that of the 3 references.

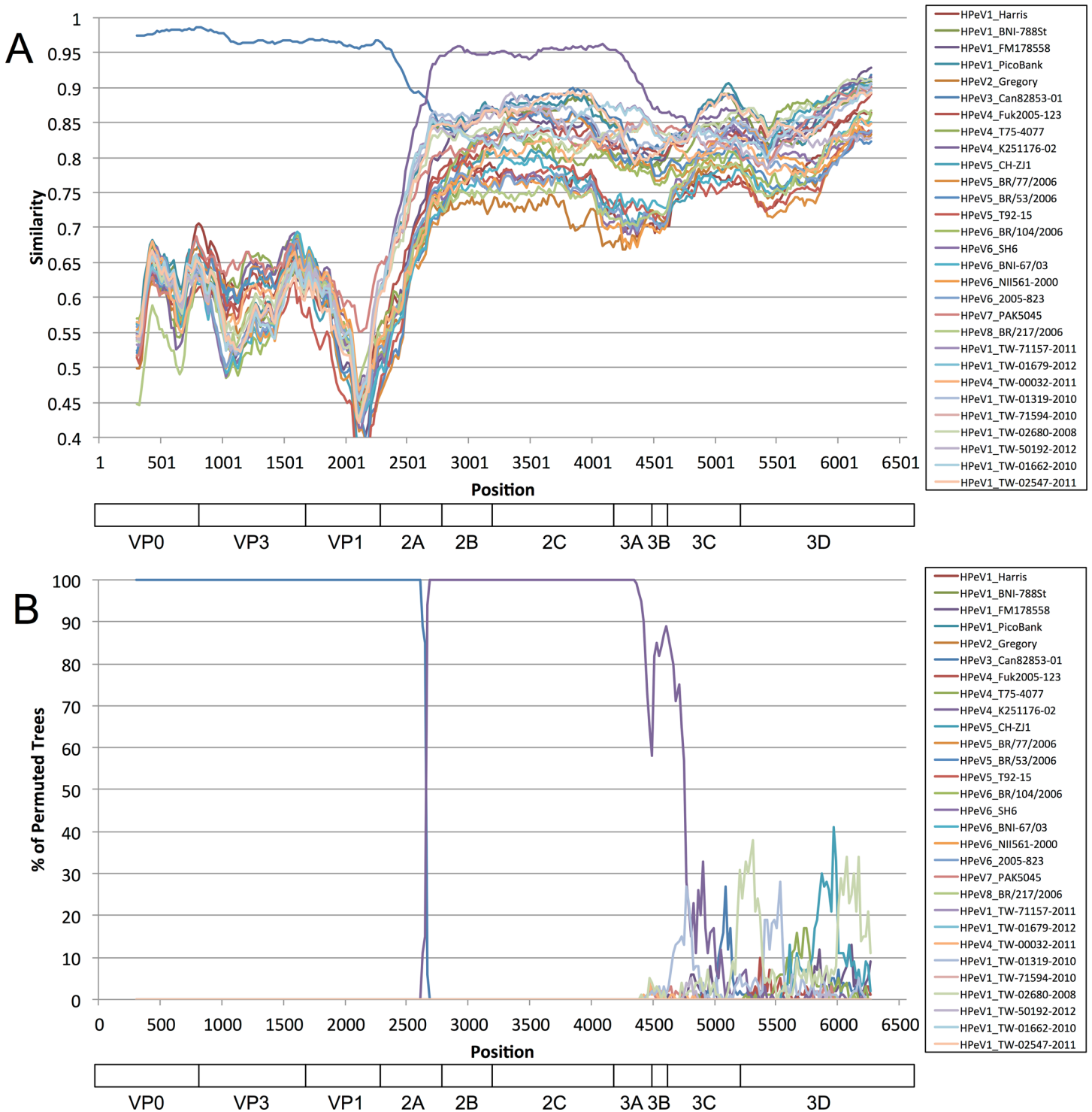
## Genetic recombination of HPeV genomes

Frequent recombination has been reported for the family of *Picornaviridae*. We performed Simplot and Bootscan analyses on the 10 TW HPeV viruses with 20 reference genomes encompassing all 8 known HPeV subtypes. The similarity plot of TW-03067-2011 (HPeV-3) resembled a HPeV-3 reference sequence (Can82853-01) in VP0, VP3, and VP1, with identity of approximately 95%–99% (Fig 5A). Thereafter, the percent identities reduced significantly to less than 88% for Can82853-01. However, an HPeV-4 strain (K251176-02) was found to have more than 96% percent identity than that of this TW HPeV-3 in the middle of the ORF spanning the C-terminal 2A to N-terminal 3C region. A Bootscan plot in Fig 5B clearly illustrates the switch over of the 2 HPeV types. Notably, the other members of Clade B also showed recombination with this HPeV-4 strain in this region, including BJ-C3174, BONN-2, VGHKS-2007, and 12 Yamagata/2008 strains (data not shown).

## Discussion

NGS technology can produce genomic sequences without the need for designing specific primers. Hence, this technology is useful for viral detection in clinical samples. A considerable number of epidemiology samples from northern Taiwan are received by CGMH for viral culturing every year; 52 544 clinical specimens were received from 2000 to 2008 [10] and 32 310 were received in the study period (2008 to 2012). CPE-positive isolates could not be identified using IFA, which might contain novel or emerging viruses that are public health risks, such as MERS-CoV, Ebola, SARS, and avian flu [33]. In addition, unknown pathogens might also exist in our culture-negative cases (approximately 60% of 32 310 clinical samples). A study reported that viral pathogens from 1406 of 2259 (approximately 60%) collected samples of patients with serious illness could also not be detected [34]; new diagnostic tests were required for those culture-negative cases.

Table 2 presents the 47 viral candidates detected using NGS, of which 35 candidates were confirmed using PCR. Among these 35 candidates, most had genome coverage higher than 80%, and only one (HRV-B4 in NGS 3) had a genome coverage of 65%. The other 12 candidates were not confirmed using PCR; their genome coverage was lower than 59%. As expected, higher genome coverage indicates that these viral genomes were well represented by the assembled contigs, and lower coverage makes confirming the existence of these viruses challenging. Although the HRV-B4 candidate in NGS 3 showed a lower coverage of 64% than that of the other 34 candidates, one contig having an extremely high depth of 10 043 with 4743 bp was categorized into this viral type, suggesting the existence of this HRV in NGS 3. By contrast, the HRV-A8 candidate in NGS 8 showed a coverage of 58%. Only one contig having an average depth of 29 with 4212 bp was categorized into this viral type. The top 3 BLAST subjects of this contig were HRV-A8, -A95, and -A45 sequences. Sequence similarities between viral pathogens



**Fig 5. Recombination analysis of HPeV genomes.** Twenty HPeV references from NCBI were retrieved in March 2015 for recombination analysis with the 10 genomes obtained in this study. (A) SimPlot and (B) Bootscan analyses of 30 HPeV ORF sequences. The query sequence was set to TW-03067-2011 of HPeV-3, which was found to resemble an HPeV-4 strain (K251176-02) from the C-terminal 2A to N-terminal 3C region.

doi:10.1371/journal.pone.0151495.g005

in mixed sample may result in a false positive rate when only considering the top one hit as the potential viral candidate.

Because 3–16 CPE-positive cases were received every year, we pooled samples for NGS analysis to reduce the sequencing cost. However, NGS analysis of mixed samples has the limitation to assign viral candidates to each isolate after BLASTN searches of quality contigs. Hence, NGS-identified viruses were confirmed and assigned to each isolate by using PCR/ELISA. All but 2 isolates (TW-01662-2010 and TW-02547-2011) were validated. An iterative mapping procedure was further performed for these two isolates that were not identified using PCR/ELISA to reveal viral genomes for detecting unknown pathogens. Although the read depths of these 2 genomes were lower than those of viral genomes from the other NGS data set, this iterative mapping process could be useful to detect unknown pathogens in clinical samples with a low viral load (or low CPE) in clinical samples. In addition, viral culturing with a low viral load negatively affects NGS experiments. Based on our experience, the time required for HPeV culture by using the RD or MRC-5 cell line is usually 3 weeks, which is much longer than 1 week for enterovirus, a common virus in regular clinical/virology laboratories. Although HPeVs have been replicated on the RD99, A549, and vero cell lines, HPeV culture is still limited by the low induction of CPE [35].

To compare current NGS pipelines including VirusFinder [36] and VirFind [37], the novelty of the proposed pipeline is to identify unknown pathogens in a mixed clinical sample, and to implement an iterative mapping for generating viral genomes. In terms of viral detection, VirusFinder could detect unknown pathogens in a single sample, but not a mixed one. The assignment of viral candidates for each sample in the proposed pipeline was by using PCR or ELISA experiments, and in VirFind by using a degenerate oligonucleotide primed RT-PCR method with multiple barcodes. The design of barcoded primers was only suitable for detecting known viruses in target samples. Moreover, an iterative mapping was implemented in this study for constructing whole genome to improve short contigs. These two pipelines did not deal with the issue of short contigs.

In this study, we used HPeV as an example to investigate the genetic diversity of viral genomes. This virus is highly infectious and causes a variety of symptoms in early childhood [38] and has atypical molecular and biological properties [13]. This study is the first to report the three genetic diversities of Taiwanese HPeV strains through various analyses, as follows. First, the phylogenetic tree in Fig 4A indicated that one HPeV-1 virus (TW-02680-2008) was highly divergent from the rest of Clade B, in which 6 nucleotide deletions were found in contrast to the other HPeV-1 viruses from the database. Although the biological consequences of this deletion are unknown, this deletion in VP0 has not been reported. Second, we suggest two cut-off values for the identification of genetic variants in a mixed sample with 1.5 gigabytes of raw reads, including a primary nucleotide composition of less than 75% and a minimum depth of 20. Based on these cut-offs, 64 genetic variants were detected from 10 HPeV strains by using NGS read mapping for revealing viral quasispecies/heterogeneity. Although additional investigations for exploring the biological consequences of these genetic variants from HPeV are required, genetic variants in the form of a viral quasispecies located in antigenic domains or drug resistant markers were similarly identified from influenza viruses. One study revealed nucleotide heterogeneity as quasispecies for characterizing influenza A/H1N1/2009 virus by using NGS read mapping [39]. Two specific amino acids (Gly172Glu and Gly239Asn) with 25% heterogeneous populations were identified, which were located on the Sa and Ca2 antigenic sites on the hemagglutinin (HA) gene, respectively. Another study reported that Arg292Lys, one of markers for N2 that reduces oseltamivir susceptibility, was a viral quasispecies [40]. Finally, recombination events among HPeV types 1, 4, 5, and 6 were frequently detected within nonstructural (NS) genes, similar to many picornaviruses, but were more restricted among HPeV-3 strains [41–43]. Our HPeV-3 strain (TW-03067-2011) was found to be recombinant with a known HPeV-4 (K251176-02). Fig

5 illustrated the switch over of the 2 HPeV types in 2A. The similarity plot of TW-03067-2011 resembled an HPeV-3 reference (Can82853-01) from VP0 to VP1, and K251176-02 from 2B to 3C. Although it was reported that this HPeV-4 strain may have a recombination event with 2 HPeV-3 strains (A308/99 and Can82853-01, as shown in Clade A of Fig 4B) in the P2–P3 region [44], our findings further indicated that HPeV-3 strains in Clade B were recombinant with this HPeV-4, showing a genomic region from the C-terminal 2A to N-terminal 3C derived from HPeV-4. In conclusion, we established the baseline information of NGS technology for viral detection and explored genetic diversities and signatures of HPeVs by the proposed NGS data analysis pipeline. The results of this study may help to improve epidemiological surveillance and investigations in Taiwan and around the world.

## Supporting Information

**S1 Fig. Deletions in VP0 gene in TW-02680-2008.** Deletions in VP0 in TW-02680-2008 are marked using a red rectangle, located at positions 33 and 34 in amino acid (or from 97 to 102 in nucleotides) based on the VP0 gene. The other HPeV-1 strains in this study and the database show “GN” at these amino acid positions.

(PDF)

**S1 File. BLAST reports of quality viral contigs with a depth of at least 20.** 364 BLASTN reports of quality viral contigs from 10 NGS experiments were summarized. The numbers of quality contigs from NGS 1 to 10 are from 7, 28, 19, 35, 48, 62, 50, 112, 2 and 1, respectively.

(XLSX)

**S2 File. Two coding sequences of TW-01662-201 and TW-02547-2011.** Two isolates (TW-01662-201 and TW-02547-2011) tested negative for PCR. An iterative mapping method was implemented to generate their coding sequences.

(FASTA)

**S3 File. Sanger sequencing bases, NGS read consensus, and depths and compositions of 64 genetic variants.** A genetic variant was defined as a nucleotide position having its primary nucleotide alphabet appearing in less than 75% of the mapped reads with read depth of at least 20. A total of 64 variants were identified using these 2 criteria. In total, 38 were found from TW-01662-2010, 9 from TW-50192-2012, 8 from TW-00032-2011 (HPeV-4), 4 from TW-02547-2011, 2 from TW-03067-2011 (HPeV-3), 1 each from TW-71157-2011, TW-01319-2010, and TW-02680-2008, and none from TW-01679-2012 and TW-71594-20110. For example, in TW-03067-201, Sanger sequencing base and NGS read consensus at position 420 in VP0 gene are “C” and “A,” respectively. This nonsynonymous substitution causes an amino acid change from Glu to Asp. Read depth at this position is 6219, including 4638 As, 1562 Cs, 15 Gs, 3 Ts, and 1 gap. In addition, 2 predicted genomes identified using iterative mapping are marked with asterisks.

(XLSX)

**S1 Table. A summary of commercial antibodies for viral screening and identification.** 5 kits for viral screening and identification were used, which contain specific antibodies for commonly seen viruses in clinical laboratory.

(DOCX)

## Acknowledgments

The authors thank help from staff of Virology Laboratory, Linkou Chang Gung Memorial Hospital, Taoyuan, Taiwan.



## Author Contributions

Conceived and designed the experiments: YNG GWC SLY CJL SRS KCT. Performed the experiments: SLY CJL KCT. Analyzed the data: YNG GWC KCT. Contributed reagents/materials/analysis tools: YNG GWC KCT. Wrote the paper: YNG GWC KCT.

## References

1. Grard G, Fair JN, Lee D, Slikas E, Steffen I, Muyembe J-J, et al. A novel rhabdovirus associated with acute hemorrhagic fever in central Africa. *PLoS Pathog*. 2012; 8: e1002924. doi: [10.1371/journal.ppat.1002924](https://doi.org/10.1371/journal.ppat.1002924) PMID: [23028323](https://pubmed.ncbi.nlm.nih.gov/23028323/)
2. Tan LV, van Doorn HR, Nghia HDT, Chau TTH, Tu LTP, de Vries M, et al. Identification of a new cyclovirus in cerebrospinal fluid of patients with acute central nervous system infections. *MBio*. 2013; 4: e00231–00213. doi: [10.1128/mBio.00231-13](https://doi.org/10.1128/mBio.00231-13) PMID: [23781068](https://pubmed.ncbi.nlm.nih.gov/23781068/)
3. Allen LZ, Ishoey T, Novotny MA, McLean JS, Lasken RS, Williamson SJ. Single virus genomics: a new tool for virus discovery. *PLoS ONE*. 2011; 6: e17722. doi: [10.1371/journal.pone.0017722](https://doi.org/10.1371/journal.pone.0017722) PMID: [21436882](https://pubmed.ncbi.nlm.nih.gov/21436882/)
4. Aw TG, Howe A, Rose JB. Metagenomic approaches for direct and cell culture evaluation of the virological quality of wastewater. *J Virol Methods*. 2014; 210C: 15–21. doi: [10.1016/j.jviromet.2014.09.017](https://doi.org/10.1016/j.jviromet.2014.09.017) PMID: [25264099](https://pubmed.ncbi.nlm.nih.gov/25264099/)
5. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 2008; 18: 821–829. doi: [10.1101/gr.074492.107](https://doi.org/10.1101/gr.074492.107) PMID: [18349386](https://pubmed.ncbi.nlm.nih.gov/18349386/)
6. Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, et al. De novo assembly and analysis of RNA-seq data. *Nat Methods*. 2010; 7: 909–912. doi: [10.1038/nmeth.1517](https://doi.org/10.1038/nmeth.1517) PMID: [20935650](https://pubmed.ncbi.nlm.nih.gov/20935650/)
7. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215: 403–410. doi: [10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: [2231712](https://pubmed.ncbi.nlm.nih.gov/2231712/)
8. Vezzi F, Cattonaro F, Policriti A. e-RGA: enhanced Reference Guided Assembly of Complex Genomes. *EMBnet.journal*. 2011; 17: 46. doi: [10.14806/ej.17.1.208](https://doi.org/10.14806/ej.17.1.208)
9. Wu TN, Tsai SF, Li SF, Lee TF, Huang TM, Wang ML, et al. Sentinel surveillance for enterovirus 71, Taiwan, 1998. *Emerging Infect Dis*. 1999; 5: 458–460. doi: [10.3201/eid0503.990321](https://doi.org/10.3201/eid0503.990321) PMID: [10341187](https://pubmed.ncbi.nlm.nih.gov/10341187/)
10. Tsao K-C, Huang C-G, Huang Y-L, Chen F-C, Huang P-N, Huang Y-C, et al. Epidemiologic features and virus isolation of enteroviruses in Northern Taiwan during 2000–2008. *J Virol Methods*. 2010; 165: 330–332. doi: [10.1016/j.jviromet.2010.03.001](https://doi.org/10.1016/j.jviromet.2010.03.001) PMID: [20214926](https://pubmed.ncbi.nlm.nih.gov/20214926/)
11. Chen G-W, Tsao K-C, Huang C-G, Gong Y-N, Chang S-C, Liu Y-C, et al. Amino acids transitioning of 2009 H1N1pdm in Taiwan from 2009 to 2011. *PLoS ONE*. 2012; 7: e45946. doi: [10.1371/journal.pone.0045946](https://doi.org/10.1371/journal.pone.0045946) PMID: [23029335](https://pubmed.ncbi.nlm.nih.gov/23029335/)
12. Kuo R-L, Yang S-L, Liu Y-C, Chen L-T, Mok C-K, Kuo S-M, et al. Influenza A/B virus detection and influenza A virus subtyping with emphasis on the novel H7N9 virus by using multiplex real-time RT-PCR. *J Virol Methods*. 2014; 208: 41–46. doi: [10.1016/j.jviromet.2014.07.020](https://doi.org/10.1016/j.jviromet.2014.07.020) PMID: [25066277](https://pubmed.ncbi.nlm.nih.gov/25066277/)
13. Shah G, Robinson JL. The particulars on parechovirus. *Can J Infect Dis Med Microbiol*. 2014; 25: 186–188. PMID: [25285119](https://pubmed.ncbi.nlm.nih.gov/25285119/)
14. Esposito S, Rahamat-Langendoen J, Ascolese B, Senatore L, Castellazzi L, Niesters HGM. Pediatric parechovirus infections. *J Clin Virol*. 2014; 60: 84–89. doi: [10.1016/j.jcv.2014.03.003](https://doi.org/10.1016/j.jcv.2014.03.003) PMID: [24690382](https://pubmed.ncbi.nlm.nih.gov/24690382/)
15. Huang Y-P, Hsieh J-Y, Wu H-S, Yang J-Y. Molecular and epidemiological study of human parechovirus infections in Taiwan, 2007–2012. *J Microbiol Immunol Infect*. 2014; doi: [10.1016/j.jmii.2014.06.013](https://doi.org/10.1016/j.jmii.2014.06.013)
16. de Crom SCM, Obihara CC, de Moor RA, Veldkamp EJM, van Furth AM, Rossen JWA. Prospective comparison of the detection rates of human enterovirus and parechovirus RT-qPCR and viral culture in different pediatric specimens. *J Clin Virol*. 2013; 58: 449–454. doi: [10.1016/j.jcv.2013.07.017](https://doi.org/10.1016/j.jcv.2013.07.017) PMID: [23973350](https://pubmed.ncbi.nlm.nih.gov/23973350/)
17. van der Sanden S, de Bruin E, Vennema H, Swanink C, Koopmans M, van der Avoort H. Prevalence of human parechovirus in the Netherlands in 2000 to 2007. *J Clin Microbiol*. 2008; 46: 2884–2889. doi: [10.1128/JCM.00168-08](https://doi.org/10.1128/JCM.00168-08) PMID: [18614653](https://pubmed.ncbi.nlm.nih.gov/18614653/)
18. Benschop K, Thomas X, Serpenti C, Molenkamp R, Wolthers K. High prevalence of human Parechovirus (HPeV) genotypes in the Amsterdam region and identification of specific HPeV variants by direct genotyping of stool samples. *J Clin Microbiol*. 2008; 46: 3965–3970. doi: [10.1128/JCM.01379-08](https://doi.org/10.1128/JCM.01379-08) PMID: [18945833](https://pubmed.ncbi.nlm.nih.gov/18945833/)
19. Malboeuf CM, Yang X, Charlebois P, Qu J, Berlin AM, Casali M, et al. Complete viral RNA genome sequencing of ultra-low copy samples by sequence-independent amplification. *Nucleic Acids Res*. 2013; 41: e13. doi: [10.1093/nar/gks794](https://doi.org/10.1093/nar/gks794) PMID: [22962364](https://pubmed.ncbi.nlm.nih.gov/22962364/)

20. Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, et al. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods*. 2010; 7: 709–715. doi: [10.1038/nmeth.1491](https://doi.org/10.1038/nmeth.1491) PMID: [20711195](https://pubmed.ncbi.nlm.nih.gov/20711195/)
21. CLC bio. Available: <http://www.clcbio.com>.
22. Deng X, Naccache SN, Ng T, Federman S, Li L, Chiu CY, et al. An ensemble strategy that significantly improves de novo assembly of microbial genomes from metagenomic next-generation sequencing data. *Nucleic Acids Res*. 2015; 43: e46. doi: [10.1093/nar/gkv002](https://doi.org/10.1093/nar/gkv002) PMID: [25586223](https://pubmed.ncbi.nlm.nih.gov/25586223/)
23. Wong S, Pabbaraju K, Pang XL, Lee BE, Fox JD. Detection of a broad range of human adenoviruses in respiratory tract samples using a sensitive multiplex real-time PCR assay. *J Med Virol*. 2008; 80: 856–865. doi: [10.1002/jmv.21136](https://doi.org/10.1002/jmv.21136) PMID: [18360899](https://pubmed.ncbi.nlm.nih.gov/18360899/)
24. Heim A, Ebnet C, Harste G, Pring-Akerblom P. Rapid and quantitative detection of human adenovirus DNA by real-time PCR. *J Med Virol*. 2003; 70: 228–239. doi: [10.1002/jmv.10382](https://doi.org/10.1002/jmv.10382) PMID: [12696109](https://pubmed.ncbi.nlm.nih.gov/12696109/)
25. Lu X, Trujillo-Lopez E, Lott L, Erdman DD. Quantitative real-time PCR assay panel for detection and type-specific identification of epidemic respiratory human adenoviruses. *J Clin Microbiol*. 2013; 51: 1089–1093. doi: [10.1128/JCM.03297-12](https://doi.org/10.1128/JCM.03297-12) PMID: [23325827](https://pubmed.ncbi.nlm.nih.gov/23325827/)
26. Nix WA, Oberste MS, Pallansch MA. Sensitive, seminested PCR amplification of VP1 sequences for direct identification of all enterovirus serotypes from original clinical specimens. *J Clin Microbiol*. 2006; 44: 2698–2704. doi: [10.1128/JCM.00542-06](https://doi.org/10.1128/JCM.00542-06) PMID: [16891480](https://pubmed.ncbi.nlm.nih.gov/16891480/)
27. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9: 357–359. doi: [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923) PMID: [22388286](https://pubmed.ncbi.nlm.nih.gov/22388286/)
28. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012; 22: 568–576. doi: [10.1101/gr.129684.111](https://doi.org/10.1101/gr.129684.111) PMID: [22300766](https://pubmed.ncbi.nlm.nih.gov/22300766/)
29. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011; 7: 539. doi: [10.1038/msb.2011.75](https://doi.org/10.1038/msb.2011.75) PMID: [21988835](https://pubmed.ncbi.nlm.nih.gov/21988835/)
30. Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol*. 2013; 30: 2725–2729. doi: [10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197) PMID: [24132122](https://pubmed.ncbi.nlm.nih.gov/24132122/)
31. Jukes T.H. and Cantor C.R. Evolution of protein molecules. In Munro HN, editor, *Mammalian Protein Metabolism*. Academic Press, New York; 1969. pp. 21–132
32. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol*. 1999; 73: 152–160. PMID: [9847317](https://pubmed.ncbi.nlm.nih.gov/9847317/)
33. Suwantarant N, Apisarnthanarak A. Risks to healthcare workers with emerging diseases: lessons from MERS-CoV, Ebola, SARS, and avian flu. *Curr Opin Infect Dis*. 2015; 28: 349–361. doi: [10.1097/QCO.000000000000183](https://doi.org/10.1097/QCO.000000000000183) PMID: [26098498](https://pubmed.ncbi.nlm.nih.gov/26098498/)
34. Jain S, Self WH, Wunderink RG, Fakhran S, Balk R, Bramley AM, et al. Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults. *N Engl J Med*. 2015; 373: 415–427. doi: [10.1056/NEJMoa1500245](https://doi.org/10.1056/NEJMoa1500245) PMID: [26172429](https://pubmed.ncbi.nlm.nih.gov/26172429/)
35. Westerhuis BM, Jonker SCM, Mattao S, Benschop KSM, Wolthers KC. Growth characteristics of human parechovirus 1 to 6 on different cell lines and cross-neutralization of human parechovirus antibodies: a comparison of the cytopathic effect and real time PCR. *Virology*. 2013; 10: 146. doi: [10.1186/1743-422X-10-146](https://doi.org/10.1186/1743-422X-10-146) PMID: [23668373](https://pubmed.ncbi.nlm.nih.gov/23668373/)
36. Wang Q, Jia P, Zhao Z. VirusFinder: software for efficient and accurate detection of viruses and their integration sites in host genomes through next generation sequencing data. *PLoS ONE*. 2013; 8: e64465. doi: [10.1371/journal.pone.0064465](https://doi.org/10.1371/journal.pone.0064465) PMID: [23717618](https://pubmed.ncbi.nlm.nih.gov/23717618/)
37. Ho T, Tzanetakis IE. Development of a virus detection and discovery pipeline using next generation sequencing. *Virology*. 2014; 471–473: 54–60. doi: [10.1016/j.virol.2014.09.019](https://doi.org/10.1016/j.virol.2014.09.019) PMID: [25461531](https://pubmed.ncbi.nlm.nih.gov/25461531/)
38. Stanway G, Joki-Korpela P, Hyypiä T. Human parechoviruses—biology and clinical significance. *Rev Med Virol*. 2000; 10: 57–69. PMID: [10654005](https://pubmed.ncbi.nlm.nih.gov/10654005/)
39. Kuroda M, Katano H, Nakajima N, Tobiume M, Aina A, Sekizuka T, et al. Characterization of quasispecies of pandemic 2009 influenza A virus (A/H1N1/2009) by de novo sequencing using a next-generation DNA sequencer. *PLoS ONE*. 2010; 5: e10256. doi: [10.1371/journal.pone.0010256](https://doi.org/10.1371/journal.pone.0010256) PMID: [20428231](https://pubmed.ncbi.nlm.nih.gov/20428231/)
40. Mok C-K, Chang S-C, Chen G-W, Lo Y-L, Chen S-J, Wu H-S, et al. Pyrosequencing reveals an oseltamivir-resistant marker in the quasispecies of avian influenza A (H7N9) virus. *J Microbiol Immunol Infect*. 2015; 48: 465–469. doi: [10.1016/j.jmii.2013.09.010](https://doi.org/10.1016/j.jmii.2013.09.010) PMID: [24388586](https://pubmed.ncbi.nlm.nih.gov/24388586/)
41. Lukashev AN. Role of recombination in evolution of enteroviruses. *Rev Med Virol*. 2005; 15: 157–167. doi: [10.1002/rmv.457](https://doi.org/10.1002/rmv.457) PMID: [15578739](https://pubmed.ncbi.nlm.nih.gov/15578739/)

42. Simmonds P, Welch J. Frequency and dynamics of recombination within different species of human enteroviruses. *J Virol.* 2006; 80: 483–493. doi: [10.1128/JVI.80.1.483-493.2006](https://doi.org/10.1128/JVI.80.1.483-493.2006) PMID: [16352572](https://pubmed.ncbi.nlm.nih.gov/16352572/)
43. Benschop KSM, de Vries M, Minnaar RP, Stanway G, van der Hoek L, Wolthers KC, et al. Comprehensive full-length sequence analyses of human parechoviruses: diversity and recombination. *J Gen Virol.* 2010; 91: 145–154. doi: [10.1099/vir.0.014670-0](https://doi.org/10.1099/vir.0.014670-0) PMID: [19759239](https://pubmed.ncbi.nlm.nih.gov/19759239/)
44. Benschop KSM, Schinkel J, Luken ME, van den Broek PJM, Beersma MFC, Menelik N, et al. Fourth human parechovirus serotype. *Emerging Infect Dis.* 2006; 12: 1572–1575. doi: [10.3201/eid1210.051647](https://doi.org/10.3201/eid1210.051647) PMID: [17176575](https://pubmed.ncbi.nlm.nih.gov/17176575/)