# Metagenomic analysis of viral genetic diversity in respiratory samples from children with severe acute respiratory infection in China

**Y. Wang[1], N. Zhu[1], Y. Li[1], R. Lu[1], H. Wang[2], G. Liu[1], X. Zou[1], Z. Xie[3] and W. Tan[1]**

1) *Key Laboratory of Medical Virology, Ministry of Health; National Institute for Viral Disease Control and Prevention,* 2) *State Key Laboratory for Infectious Disease Prevention and Control and National Institute for Communicable Disease Control and Prevention, Chinese Centre for Disease Control and Prevention and* 3) *Key Laboratory of Major Diseases in Children and National Key Discipline of Paediatrics Capital Medical University, Ministry of Education, Beijing Pediatric Research Institute, Beijing Children's Hospital, Capital Medical University, Beijing, China*

## Abstract

Severe acute respiratory infection (SARI) in children is thought to be mainly caused by infection with various viruses, some of which have been well characterized; however, analyses of respiratory tract viromes among children with SARI versus those without are limited. In this study, nasopharyngeal swabs from children with and without SARI (135 versus 15) were collected in China between 2008 and 2010 and subjected to multiplex metagenomic analyses using a next-generation sequencing platform. The results show that members of the *Paramyxoviridae*, *Coronaviridae*, *Parvoviridae*, *Orthomyxoviridae*, *Picornaviridae*, *Anelloviridae* and *Adenoviridae* families represented the most abundant species identified (>50% genome coverage) in the respiratory tracts of children with SARI. The viral population found in the respiratory tracts of children without SARI was less diverse and mainly dominated by the *Anelloviridae* family with only a small proportion of common epidemic respiratory viruses. Several almost complete viral genomes were assembled, and the genetic diversity was determined among several samples based on next-generation sequencing. This research provides comprehensive mapping of the viromes of children with SARI and indicates high heterogeneity of known viruses present in the childhood respiratory tract, which may benefit the detection and prevention of respiratory disease.

**Corresponding author:** W. Tan, Key Laboratory of Medical Virology, Ministry of Health, National Institute for Viral Disease Control and Prevention, China CDC, 155 Changbai Road, ChangPing District, Beijing102206, China
**E-mail:** tanwj28@163.com

## Introduction

Severe acute respiratory infection (SARI) is responsible for a significant proportion of the paediatric disease burden worldwide, especially in China [1]. Previous estimates indicated that 1.9 million children worldwide die each year from acute respiratory illnesses, many of which are SARI [2]. Furthermore, most of the cases of SARI in China (74.7% or 94.29%) resulted from viral pathogens, according to the latest report [3]. It is important to catalogue viruses present in the respiratory tracts of children as comprehensively as possible. However, it is difficult to detect all viral pathogens and make a definite diagnosis by routine detection methods because of the variety of respiratory viruses and genetic variations.

In recent years, high-throughput next-generation sequencing (NGS) technology has become a powerful tool in pathogen detection compared with traditional methods, because it allows for the detection of pathogens without any advance genetic information. To date, NGS has been widely used for the characterization of community viral infections and novel

pathogen identification in numerous hosts, including humans [4,5], pigs [6], ferrets [7], bats [8,9], cats [10] and rats [11]; it has also been used in locations such as vineyards [12] and in freshwater [13].

The application of NGS to children's respiratory tracts was expected to provide comprehensive information about the respiratory tract virome and allow characterization of specific viruses. Results from investigations of the virome of children's respiratory tracts have been reported by several groups [14–17]. Metagenomic detection of viral pathogens in nasal and faecal specimens in Japan indicated that the unbiased high-throughput sequencing approach is useful for directly detecting pathogenic viruses [18]; however, genome analysis was seldom available, due to the majority of reads (> 90%) derived from the host genome. Because of the high proportion of host genomic and bacterial sequences, viral reads were rarely detected in respiratory samples by high-throughput sequencing, or only a small proportion (<1%) of generated reads were assigned to known viruses. Most studies of the respiratory tract virome were conducted using routine molecular detection methods [19,20]. For example, one recent report about viral profiles in China showed that piconaviruses and para-myxoviruses are the most common viral pathogens among children with SARI using a commercial assay for 18 respiratory viral targets based on multiplex PCR [20]. Comprehensive and detailed metagenomic analyses of respiratory tracts of children with SARI and matched control groups have not yet been reported. Therefore, we conducted a viral metagenomic analysis and attempted to provide a complete picture of the viral content and diversity in the respiratory tracts of children with SARI.

## Materials and Methods

### Study population and sample collection
We randomly enrolled 135 paediatric patients <6 years of age in Beijing Children's Hospital from 2008 to 2010 who met the following SARI case definition [21]: reported temperature (rectal or axillary) of $\geq 38°C$ and one of the following signs and symptoms: cough, sore throat, tachypnoea, abnormal breath sounds, dyspnoea; and chest pain or chest radiographic documentation of parenchymal lung abnormalities. In addition, 15 respiratory samples from children without respiratory symptoms were randomly collected as control samples. All the samples were collected between May 2008 and October 2010, the mean age of SARI and control group were 1.18 and 4.26 years, respectively.

Sample information, including the children's age and sex, sample collection date, and respiratory symptoms, is presented

in the Supplementary material (Table S1). This study was performed in strict accordance with human subject protection guidance set out by the Research Ethics Committee of Beijing Children's Hospital and the Institutional Review Board at the China CDC. Written informed consent was obtained on the participants' behalf from their parents or guardians.

### Viral enrichment and random amplification
Respiratory samples (nasopharyngeal swabs) were prepared as detailed in a previously published protocol [8,15] with some modifications. In brief, 135 samples from children with SARI were processed into nine pools of 15 samples each. The viral transport medium was filtered through a 0.22-μm filter, and the filtrates were ultra-centrifuged at 30,000 rpm (154,000 g) in an SW41 rotor (Beckman, Miami, FL, USA) for 3 h to pellet the viral particles. The pellets were treated with a cocktail of nucleases (DNase, RNase and nuclease) to remove any remaining extracellular nucleic acids. Nucleic acid extraction was performed using the QIAamp MinElute Virus Spin Kit. Extracted nucleic acids (DNA and RNA) were amplified by sequence-independent single primer amplification. Meanwhile, 15 respiratory samples from the control group were divided into four pools of three to five samples each and processed as above.

### Metagenomic analysis based on NGS
The amplified DNA was used as a template for Illumina sequencing; the sequencing procedure was conducted according to the Illumina Hiseq 2500 sequencing protocol used by the Berry Genomics Company (Beijing, China). Metagenomic analysis was performed as below, firstly, paired-end reads (2 × 125-bp reads) from the Hiseq 2500 instrument were quality-trimmed, completely identical sequences were collapsed into a single representative sequence to minimize the number of reads that needed to be analysed, then low-complexity reads and those <50 bases long were removed using a FASTX-Toolkit. Second, human genome sequences were identified for removal by aligning sequences to the human genome, using SOAP ALIGNER with the following parameters: 70% coverage and 90% identify. Third, non-human sequences were assembled into contigs by CLC GENOMIC WORKBENCH and SOAP DE NO, and aligned to the metagenomic database comprising all virus sequences in the NCBI nt (non-redundant nucleotide, E value less <10e$^{-03}$) and nr databases (non-redundant protein, E value <1000). The highest scoring hit was assigned as the closest homologue.

### Phylogenetic analysis
To analyse genetic variation of the predominant respiratory viruses detected in this study, nucleotide sequences were compared with entries in the GenBank database. Nucleotide

sequence alignment was conducted via the MAFFT programme, version 5. Phylogenetic and molecular evolutionary analyses were conducted using the neighbour-joining method using MEGA 5.0 with a bootstrap value of 1000.

### Nucleotide sequence accession numbers

All complete or partial genome sequences were logged in the GenBank (KR607978–KR607996). All metagenomic sequencing reads were submitted to the NCBI Sequence Read Archive (Accession numbers: SRR2010685, SRR2010686, SRR2040553 and SRR2040557). Data were analysed by the chi-squared test using SAS software version 9.2. Values of $p < 0.05$ were considered statistically significant.

## Results

### Overview of sequencing data

Multiplexed respiratory samples selected from children with SARI were processed into nine pools (A to I pools) and analysed by Illumina Hiseq 2500 paired-end sequencing. The total number of reads per sample pool ranged from 1 566 014 to 11 291 524, and the proportion of reads related to viruses (without bacteriophage) compared with all reads in each sample pool ranged from 2.8% to 64.7% (average rate 17.2%; Table 1). Meanwhile, 15 respiratory samples from the control group were grouped into four pools (A1, B1, C1 and D1 pools), and the proportion of virus-specific reads in a sample pool ranged from 0.1% to 2.9% (average rate 1.1%; Table 2). Significant differences were observed regarding proportion of virus-specific reads between SARI and control group ($p < 0.001$).

### Respiratory tract virome analyses from children with and without SARI

Viral metagenomic analyses revealed a variety of respiratory viruses in the respiratory tracts of children with SARI; of these, most of the viral reads could be classified into seven known virus families: *Paramyxoviridae*, *Coronaviridae*, *Parvoviridae*, *Anelloviridae*, *Orthomyxoviridae*, *Picornaviridae* and *Adenoviridae*. The detailed results are summarized in Fig. 1, which shows that the relative abundance of reads of different viruses varied between respiratory sample pools. The most highly represented respiratory viruses were, in order of sequence read abundance, human respiratory syncytial virus (HRSV) (48.1% of all non-

**TABLE 1. Reads numbers derived from respiratory sample pools of children with severe acute respiratory infection, listed according to their similarities to virus group**

| Family | Virus genus/Read numbers | Respiratory sample pools | | | | | | | | |
| | | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|---|
| Paramyxoviridae | HRSV/6,252,381 | 5 134 888 | 315 136 | 242 | 12 045 | 2231 | 484 118 | 285 844 | 5174 | 12 703 |
| | HPIV1/10,877 | 117 | 6739 | 3685 | | 332 | | 4 | | |
| | HPIV3/249,934 | 150 | 134 031 | | 114 785 | 72 | | 8 | 332 | 556 |
| | HMPV/6,255 | 6185 | | | | 70 | | | | |
| | Measles/2,705 | | | | 1247 | 30 | | | 1426 | 2 |
| Coronaviridae | HCoV-OC43/1,605,644 | 21 013 | 554 316 | 373 | 21 061 | 997 171 | | 624 | 10 500 | 586 |
| | HCoV-229E/29,216 | 393 | 3007 | 16 856 | 7584 | 565 | 789 | 18 | | 4 |
| | HCoV-NL63/1,595 | | | 1385 | | | 196 | 6 | 8 | |
| | HCoV-HKU1/1,529,647 | 4012 | 1 486 766 | | 38 703 | 166 | | | | |
| Orthomyxoviridae | Influenza A/34,674 | 100 | 1068 | 18 | 11 914 | 9646 | 687 | 10 | | 11 231 |
| | Influenza B/3,,985 | | 397 | | | 2832 | | 5 | | 751 |
| | Influenza C/372 | | | | | 354 | | | | 18 |
| Parvoviridae | HBoV/953,144 | 50 633 | 34 954 | 27 635 | 712 588 | | | 88 089 | 37 032 | 2213 |
| Picornaviridae | HRV-A/26,130 | 167 | 2677 | 14 438 | 7129 | 364 | 289 | 236 | 600 | 230 |
| | HRV-B/3,721 | 1435 | 365 | | 1572 | | | 112 | 203 | 34 |
| | HRV-C/5,739 | 2379 | 1394 | | 1614 | 118 | 76 | 102 | 48 | 8 |
| | Human poliovirus 2/35 | | | | | 35 | | | | |
| Adenoviridae | HAdV-B/10,010 | 98 | 651 | 1875 | 2547 | | 286 | 47 | 4432 | 74 |
| | HAdV-C/1,999 | 144 | 485 | | 368 | | 119 | | 857 | 26 |
| | Other HAdVs/384 | | 72 | | 158 | | 67 | | 86 | 1 |
| Anelloviridae | TTVs/1,976,190 | 822 | | 178 249 | 321 129 | | 335 | 555 687 | 229 716 | 690 252 |
| | TTMVs/290,732 | 506 | | 7514 | 2085 | | 2964 | 9395 | 174 099 | 94 169 |
| Other viruses | HBV/56 | 56 | | | | | | | | |
| | HCV/1,372 | 670 | 71 | | 145 | | | 330 | | 156 |
| | HPVs/2,153 | 867 | 158 | | | | 1128 | | | |
| | HHVs/8,240 | 1185 | 95 | 875 | 151 | | 3483 | 170 | | 2281 |
| | Bovine viral diarrhoea virus/657 | 128 | | | 364 | 165 | | | | |
| | Rodent Torque teno virus/420 | 25 | 347 | | | | 48 | | | |
| Number of generated reads per sample pool | | 8 687 826 | 9 521 164 | 7 963 520 | 7 053 768 | 1 566 014 | 8 158 998 | 10 751 298 | 11 291 524 | 10 483 676 |
| Proportion of virus-related reads compared with all reads (excluding bacteriophages) | | 60.1% | 26.7% | 2.8% | 17.7% | 64.7% | 6.0% | 8.7% | 4.1% | 7.8% |

Note: Human genome, bacteria and bacteriophages have been excluded.
HAdV, human adenoviruses; HBoV, human bocavirus; HCoV, human coronaviruses; HRV, human rhinovirus; HRSV, human respiratory syncytial virus; TTV, Torque teno virus; TTMV, Torque teno mini virus; HPIV, human parainfluenza viruses.

**TABLE 2. Reads numbers derived from respiratory sample pools of children without respiratory symptoms, listed according to their similarities to virus group**

| Family | Virus genus/Reads numbers | Respiratory sample pool | | | |
|---|---|---|---|---|---|
| | | A1 | B1 | C1 | D1 |
| Anelloviridae | TTV/11,752 | 533 | 10 904 | 149 | 166 |
| | TTV1/8,804 | 3539 | 1029 | | 4236 |
| | TTV3/39,890 | 6967 | 29 835 | 706 | 2382 |
| | TTV4/1183 | 328 | 722 | 52 | 81 |
| | TTV6/49,367 | 1011 | 47 749 | 149 | 458 |
| | TTV7/22,668 | 4769 | 17 225 | 200 | 474 |
| | TTV8/24,919 | 12 076 | 7557 | 1677 | 3609 |
| | TTV10/43,043 | 5960 | 31 799 | 1798 | 3486 |
| | TTV12/10,915 | 4347 | 2704 | 994 | 2870 |
| | TTV15/11,842 | 5049 | 4171 | 784 | 1838 |
| | TTV16/31,711 | 5768 | 21 687 | 587 | 3669 |
| | TTV19/158,780 | 21 875 | 132 129 | 1009 | 3767 |
| | TTV28/6532 | 524 | 5495 | 88 | 425 |
| | TTMV/252 | 160 | 61 | 31 | |
| | TTMV3/43 | 7 | 32 | 4 | |
| | TTMV5/880 | 278 | 84 | 517 | 1 |
| | TTMV7/81 | 2 | 27 | 50 | 2 |
| | TTMV9/97 | 55 | 36 | 6 | |
| | TTMDV1/809 | 312 | 379 | 118 | |
| | TTMDV2/732 | 295 | 381 | 45 | 11 |
| | TTV-like mini virus/166 | 102 | 52 | | 12 |
| | Small anellovirus 1/388 | 112 | 244 | 26 | 6 |
| | Small anellovirus 2/580 | 197 | 340 | 34 | 9 |
| Others | HAdVs/1735 | 8 | 908 | 767 | 52 |
| | HRVs/121 | 32 | 44 | 43 | 2 |
| | HRSV/16 | 16 | | | |
| | HHVs/10,769 | 3026 | 121 | 87 | 7535 |
| | Influenza virus A/3 | 3 | | | |
| | HCoV-NL63/934 | | 934 | | |
| | HBoV/109 | | 109 | | |
| | HPIVs/33,350 | | 33 350 | | |
| | Flock house virus/4,301 | 2324 | 190 | | 1787 |
| | Porcine stool-associated circular virus/12 | 12 | | | |
| Number of generated reads per sample pool | | 10 959 072 | 11 992 160 | 10 683 380 | 9 753 706 |
| Proportion of virus-related reads compared with all reads | | 0.7% | 2.9% | 0.1% | 0.4% |

HAdV, human adenoviruses; HBoV, human bocavirus; HCoV, human coronaviruses; HRV, human rhinovirus; HRSV, human respiratory syncytial virus; TTV, Torque teno virus; TTMV, Torque teno mini virus; TTMDV, Torque teno midi virus; HPIV, human parainfluenza viruses.

bacteriophage virus reads), human coronaviruses (HCoVs) (24.3%), Torque teno virus (TTVs) (17.4%), human bocavirus (HBoV) (7.3%), human parainfluenza viruses (HPIV) (2.0%), human rhinovirus (HRV) (0.3%), influenza viruses (0.3%) and human adenoviruses (HAdV) (0.1%). These virus species above together comprised almost 99% of the viral contigs or reads in all sample pools (Table 1). However, some sequences matching with other eukaryotic viruses were also detected within the respiratory tract, but at much lower read abundance. These included human metapneumovirus, measles virus, hepatitis viruses B and C, human papillomavirus, human poliovirus 2 and influenza viruses B and C (Table 1). The poliovirus sequences identified were likely to be derived from the oral vaccine (Sabin type 2 vaccine strain) and hepatitis viruses B and C sequences might come from chronic hepatitis B or C carrier children.

The viral composition in samples from the respiratory tracts of children without SARI was less complex than that in children with SARI, with 89.3% of all non-bacteriophage virus reads belonging to the *Anelloviridae* family (Fig. 1). The reads mainly comprised TTV, Torque teno mini virus (TTMV), Torque teno midi virus (TTMDV) and small anellovirus (SAV). Moreover, numerous different genotypes of TTV, TTMV and TTMDV were found in the respiratory tracts of children without SARI (Table 2). Second to the *Anelloviridae* family, several epidemic respiratory viruses were detected, including human herpesviruses (HHVs), HAdV, HRV and influenza virus A. The obvious differences in reads number and proportion between the SARI and control groups indicated that the viral population was dominated by common respiratory viruses in the SARI pool and TTV-related viruses in control group.

**Comparison for the causes of SARI**
Whether the viral reads identified reflect virus shedding in acute respiratory infections or the residual genomic segment will require the full-genome analysis of individual virus through
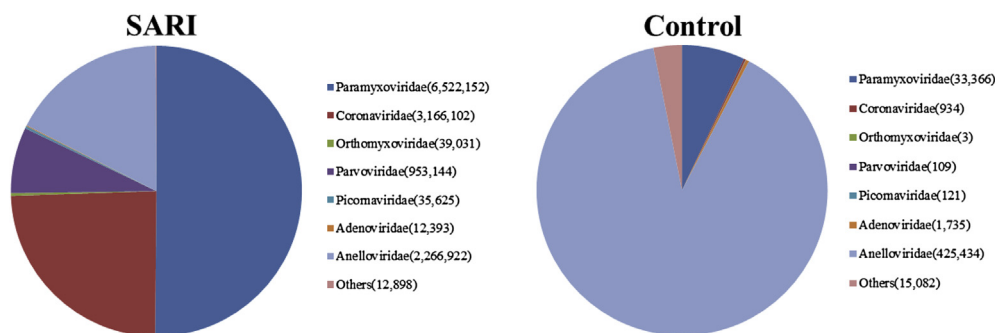


**FIG. 1.** An overview of class distribution of viral content from children with severe acute respiratory infection (SARI) and control group. There were seven main families in the respiratory tract of children with SARI, including Paramyxoviridae, Coronaviridae, Parvoviridae, Orthomyxoviridae, Picornaviridae, Adenoviridae and Anelloviridae.

conventional detection methods. Here, we separate all the viruses with >50% genome coverage and made the comparison between children with SARI and the control group. The high degree of virus diversity with distinct respiratory viruses seen here in children with SARI is greater than that of the control group. As seen in Table 3, the average number of common respiratory viruses per sample pool was 3.4 (from one to seven) for children with SARI and was 0.5 for control group. According to comparative analysis, the viruses with >50% coverage rate, including HRSV, HCoVs (HCoV-OC43, 229E and HKU1), HBoV, HPIVs, influenza virus A and HAdV, may be the main causes for SARI in Beijing's children.

### TTV and other viruses

In this research, diverse types of family *Anelloviridae* were found in respiratory samples of children with SARI and the control group, including TTVs, TTMVs, TTMDVs and SAVs. We collected all the members of family *Anelloviridae* with >50% genome coverage and analysed the relationship between acute respiratory infection and TTVs, the viruses related to the family *Anelloviridae* included TTV, TTV1, TTV3, TTV6, TTV7, TTV8, TTV10, TTV12, TTV15, TTV16, TTV19, TTV27, TTV28, TTMV, TTMV8 and SAV1 (see Supplementary material, Fig. S1), and furthermore most of the viruses with >50% genome coverage belong to TTVs (see Supplementary material, Table S2). As to the types of TTVs, there was little difference between children with SARI and the control group. In addition, TTVs were detected in sample pools from eight of nine children with SARI and four of four from the control group (see Supplementary material, Table S2), indicating that there might be no direct link between acute respiratory infection and TTVs; and children might constitute an important reservoir of TTVs.

Interestingly, several sequences related to zoonotic viruses were also present in the respiratory tracts of these children, albeit at a lower frequency, including flock house virus, porcine stool-associated circular virus, bovine diarrhoea virus and rodent TTV (Tables 1 and 2).

In addition to the reads matching the above-mentioned human and eukaryotic viruses, a large proportion of phage-related reads were detected in the respiratory tract (data not shown), mainly including *Enterobacteria* phage, *Streptococcus* phage, *Lactococcus* phage, *Bacillus* phage and *Erwinia* phage. *Haemophilus* phage, *Salmonella* phage and *Yersinia* phage were also observed on several occasions. The variety of phage is an important part of the respiratory virome and may be associated with bacterial infections.

### Sequencing depth and coverage

To eliminate the possible bias of random amplification and NGS methods, four representative viruses were used to compare the genome coverage and sequencing depth, including HRSV (an RNA virus) from pool A, HBoV (a DNA virus) from pool B, TTMV8 (a circular DNA virus) from pool C and influenza A virus (a segment RNA virus) from pool E. Reference sequences of these three representative viruses were simultaneously used in the CLC Genomic Workbench mapping analysis to determine the genome coverage and sequencing depth. As in Fig. 2, the genome coverages for the four representative viruses were 97.49% (HRSV), 100.00% (HBoV1), 99.48% (TTMV8) and 80.37% (influenza A virus, segment 8); the average depths of sequencing were 77.78 (HRSV), 12 758.93 (HBoV1), 82.50 (TTMV8) and 33.03 (influenza A virus, segment 8), respectively. The random distribution of reads on the representative virus genomes indicated that most of the respiratory virus genomes

**TABLE 3. Summary of viruses detected in the respiratory samples with over 50% genome coverages, in order of decreasing coverage rates, members of the family Anelloviridae were not included**

| Pool | Children with severe acute respiratory infection | | | | | | | | | Control groups (without SARI) | | | |
| | A | B | C | D | E | F | G | H | I | A1 | B1 | C1 | D1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Virus | HBoV (100%) HRSV (99.9%) OC43 (91.3%) | HBoV (98.1%) HPIV3 (96.9%) HKU1 (92.5%) HRSV (88.0%) | 229E (99.7%) HPIV1 (99.6%) HBoV (70.6%) | HBoV (100%) HRSV (99.5%) HKU1 (99.5%) HPIV3 (99.5%) 229E (98.1%) OC43 (94.8%) H1N1 (56.2%) | HRSV (99.5%) Influ virus B (96.3%) H1N1 (79.5%) Influ virus C (75.2%) OC43 (71.4%) | HRSV (99.2%) | HBoV (100%) HRSV (99.4%) | HBoV (100%) OC43 (99.9%) HAdV7 (91.8%) HRSV (86.5%) | HBoV (96.2%) H1N1 (59.1%) | | HAdV2 (51.5%) | HHV7 (62.8%) | |

Abbreviations: HAdV, human adenoviruses; HBoV, human bocavirus; HCoV, human coronaviruses; HRV, human rhinovirus; HRSV, human respiratory syncytial virus; TTV, Torque teno virus; HPIV, human parainfluenza viruses; OC43, HCoV-OC43; HKU1, HCoV-HKU1; 229E, HCoV-229E; H1N1, influenza virus A subtype H1N1; Influ virus B/C, influenza virus B/C.
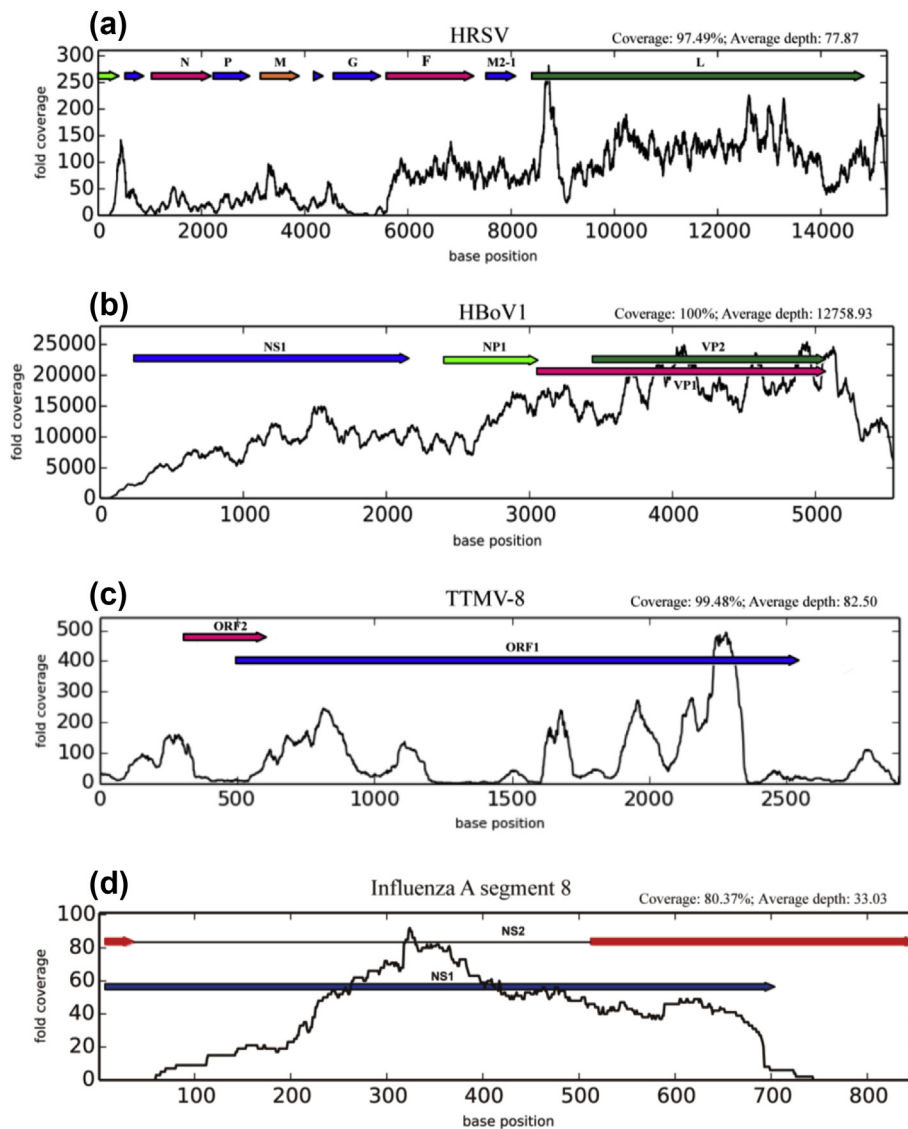
FIG. 2. Viral genome coverage and sequencing depth analysis. The three representatives included (a) human respiratory syncytial virus (HRSV), (b) human bocavirus 1 (HBoV1), (c) Torque teno mini virus 8 (TTMV-8), (d) influenza A virus segment 8. The main genes of every virus are indicated in the corresponding figures.

could be successfully obtained through random amplification and NGS methods in this study.

## Discussion

Here, we conducted a metagenomic analysis of the virome of both children with SARI and a control group (children without SARI). This study is the first to perform a full, detailed analysis of the virome of respiratory tract samples among children with SARI and a cohort group. Several metagenomic analyses of the respiratory virome of children have been conducted. Lysholm

*et al.* [15] analysed the viral microbiome in patients with severe lower respiratory tract infections and found three major families: *Paramyxoviridae*, *Picornaviridae* and *Orthomyxoviridae*. Additionally, the predominant genera identified included HRSV, influenza virus A and HRV, whereas influenza A and HRV represented a small proportion of viral reads in the present analysis. Zoll *et al.* [22] reported that HRV, TTV and TTMV were frequently detected in children and that viral populations were less complicated than those found in children from Beijing. The only viral metagenomic research on respiratory samples was reported by Yang *et al.* [17] in China, in which only 0.05% of the sequences were virally derived. These differences

may be the result of differences in the respiratory samples of children, sample collection, location and time, because the residual genome of host and bacteria will greatly affect the later metagenomic analysis. The respiratory tracts of children were likely to act as mixing vessels in which diverse viruses could be detected, such as common respiratory viruses, zoonotic pathogens, and some viruses associated with non-respiratory viral infections.

The main pathogens related to SARI were not restricted to one or two viral pathogens, but a diverse viral community in the respiratory tract [23]. Metagenomic analysis in this study demonstrated that the vast majority of viruses identified belonged to several abundant families. Viral composition differed both between and within children with SARI and the control group. By comparison of viruses with high coverage rate between children with SARI and the control group, HRSV, HCoVs (HCoV-229E, HCoV-OC43, HCoV-HKU1), HBoV, influenza virus A (H1N1), HAdVs and HPIVs may be the main causes for SARI in Beijing's children. These data agree with our previous investigation on paediatric patients with SARI in China [19]. Regarding the difference in the virome of children with SARI versus those without, the fraction of children with SARI who were shedding any one of the common respiratory viruses was much greater than that of children without SARI, except for the *Anelloviridae* family.

Differences between SARI and control groups were most striking in the proportion and composition of the main respiratory viruses, which may be associated with the SARI. As expected, the virome of the SARI group was more complex than that of the control group—on average the samples pooled from the SARI group contained about six- to seven-fold more viral pathogens with >50% coverage than the samples pooled from the control group. Previous research about the virome in febrile and afebrile groups came to a similar conclusion [24]. The nasopharyngeal swabs samples from febrile children contained more viral sequences than those from afebrile children. Whereas some viruses were detected in both SARI and control groups, HRV, HRSV, HAdV and HBoV, which are known to be associated with SARI, were particular interest. We suggest that viral detection could be asymptomatic and occasionally prolonged, just as the previous report indicated that the interpretation of positive PCR for some viruses (especially for HRV and HBoV) may be confounded, particularly in young children and those who live with them [25].

In this study, diverse members of the *Anelloviridae* family were found in respiratory samples of children with SARI and children in the control group. TTV was mainly found in blood; however, in this study, TTV and TTMV represented the vast majority of viral reads in several pools, especially in children without respiratory symptoms. The relationship between respiratory infection and TTV remains unknown. Diverse types of TTV and TTMV found in respiratory samples of children may be related to the body's immune status [26,27]; some reports have described that the TTV load was significantly increased when the immune status was in decline [28]. No association was found between the presence of TTV and SARI; all the observations indicated that children might constitute an important reservoir of TTVs. In addition, some reads matching with zoonotic pathogens were detected in the respiratory tracts of both children with SARI and controls, such as flock house virus, porcine stool-associated circular virus, bovine diarrhoea virus and rodent TTV. The source of these zoonotic pathogens found in the current study may be the environment, diet or a different source [29]. Further sampling should determine if the prevalence of such diverse zoonotic pathogens and viral loads are a general phenomenon in children with SARI and controls.

Based on the metagenomic analysis of the virome, the proportion of viral reads reached 0.1% to 64.7% in this study. Considering that all sample pools were subjected to the same sample preparation process, such large differences in viral proportions are not likely to be the result of errors in manual operations. In addition, by performing sample pre-preparation and viral purification, most human and bacterial genome reads in respiratory samples should be eliminated, with the exception of bacteriophages. The proportion of bacteriophage among all generated reads per sample pool not only influenced the viral composition, but also revealed the presence of other pathogens in respiratory samples, such as various bacteria. Previous metagenomic analyses of respiratory samples from children with lower and upper respiratory tract infections have indicated that the cases associated with bacterial infection accounted for a small proportion in the SARI cases [30], especially for those with unexplained fever. More studies on the relationship between respiratory diseases and the diversity of bacteriophages are needed.

Regarding the amplification bias of DNA and RNA viruses, there was no obvious difference between the sequencing depth and coverage for different viruses as the results showed. However, due to the limitation of sample collection and children enrolled in this study, there are still some problems in the experimental design, which may influence the results, including the sample size and the mean age for each group. There are 135 children with SARI and 15 children without SARI involved in this study. The mean ages of the SARI and control groups were 1.2 versus 4.3 years, respectively. These obvious differences may have some impact on the final result. Considering the previous research about aetiology [23,24,30], there was no obvious difference among the viral pathogens for children <6 years old, which could be treated as an independent group

for metagenomic analysis. Regarding the sampling size per group, there was a visible difference in the viral diversity and main viral population with over 50% genome coverage rate between the SARI and control groups, which is sufficient to reach a preliminary conclusion and the corresponding influence should be limited. More respiratory samples were needed to provide more detailed information about the virome characteristics of the children with SARI.

In summary, we conducted a metagenomic analysis of the virome in both children with SARI and control children. Comparisons of the sequence reads provided a global taxonomic distribution of viral reads for each sample pool, highlighting differences both within and between children with SARI and those without. This study provides the first preliminary understanding of the virome of both children with SARI and asymptomatic children, which will benefit the global surveillance, prevention and control of acute respiratory infection.

## Acknowledgements

## Transparency declaration

The authors declare that they have no conflict of interest.

## Appendix A. Supplementary data

Additional Supporting Information may be found in the online version of this article can be found at http://dx.doi.org/10.1016/j.cmi.2016.01.006.

## References

[1] Nair H, Nokes DJ, Gessner BD, Dherani M, Madhi SA, Singleton RJ, et al. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. Lancet 2010;375:1545–55.

[2] Klig JE, Shah NB. Office pediatrics: current issues in lower respiratory infections in children. Curr Opin Pediatr 2005;17:111–8.

[3] Wei L, Liu W, Zhang XA, Liu EM, Wo Y, Cowling BJ, et al. Detection of viral and bacterial pathogens in hospitalized children with acute respiratory illnesses, Chongqing, 2009–2013. Medicine 2015;94: e742.

[4] Finkbeiner SR, Allred AF, Tarr PI, Klein EJ, Kirkwood CD, Wang D. Metagenomic analysis of human diarrhea: viral detection and discovery. PLoS Pathog 2008;4:e1000011.

[5] Victoria JG, Kapoor A, Li LL, Blinkova O, Slikas B, Wang CL, et al. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. J Virol 2009;83:4642–51.

[6] Shan T, Li L, Simmonds P, Wang C, Moeser A, Delwart E. The fecal virome of pigs on a high-density farm. J Virol 2011;85:11697–708.

[7] Lin Z, Farooqui A, Li G, Wong GK, Mason AL, Banner D, et al. Next-generation sequencing and bioinformatic approaches to detect and analyze influenza virus in ferrets. J Infect Dev Countr 2014;8: 498–509.

[8] Donaldson EF, Haskew AN, Gates JE, Huynh J, Moore CJ, Frieman MB. Metagenomic analysis of the viromes of three north american bat species: viral diversity among different bat species that share a common habitat. J Virol 2010;84:13004–18.

[9] Ge X, Li Y, Yang X, Zhang H, Zhou P, Zhang Y, et al. Metagenomic analysis of viruses from the bat fecal samples reveals many novel viruses in insectivorous bats in China. J Virol 2012;JVI:0667111.

[10] Ng TF, Mesquita JR, Nascimento MS, Kondov NO, Wong W, Reuter G, et al. Feline fecal virome reveals novel and prevalent enteric viruses. Vet Microbiol 2014;171:102–11.

[11] Sachsenröder J, Braun A, Machnowska P, Ng TF, Deng X, Guenther S, et al. Metagenomic identification of novel enteric viruses in urban wild rats and genome characterization of a group a rotavirus. J Gen Virol 2014;95:2734–47.

[12] Coetzee B, Freeborough M-J, Maree HJ, Celton J-M, Rees DJG, Burger JT. Deep sequencing analysis of viruses infecting grapevines: virome of a vineyard. Virology 2010;400:157–63.

[13] Djikeng A, Kuzmickas R, Anderson NG, Spiro DJ. Metagenomic analysis of rna viruses in a fresh water lake. PLoS One 2009;4:e7264.

[14] Bogaert D, Keijser B, Huse S, Rossen J, Veenhoven R, van Gils E, et al. Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. PLoS One 2011;6:e17035.

[15] Lysholm F, Wetterbom A, Lindau C, Darban H, Bjerkner A, Fahlander K, et al. Characterization of the viral microbiome in patients with severe lower respiratory tract infections, using metagenomic sequencing. PLoS One 2012;7:e30875.

[16] Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, Silva J, et al. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. PLoS One 2009;4: e7370.

[17] Yang J, Yang F, Ren L, Xiong Z, Wu Z, Dong J, et al. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. J Clin Microbiol 2011;49:3463–9.

[18] Nakamura S, Yang CS, Sakon N, Ueda M, Tougan T, Yamashita A, et al. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. PLoS One 2009;4:e4219.

[19] Lu R, Yu X, Wang W, Duan X, Zhang L, Zhou W, et al. Characterization of human coronavirus etiology in chinese adults with acute upper respiratory tract infection by real-time RT-PCR assays. PLoS One 2012;7:e38638.

[20] Zhang C, Zhu N, Xie Z, Lu R, He B, Liu C, et al. Viral etiology and clinical profiles of children with severe acute respiratory infections in China. PLoS One 2013;8:e72606.

[21] Wharton M, Chorba TL, Vogt RL, Morse DL, Buehler JW. Case definitions for public health surveillance. MMWR Recomm Rep 1990;39: 1–43.

[22] Zoll J, Rahamat-Langendoen J, Ahout I, de Jonge MI, Jans J, Huijnen MA, et al. Direct multiplexed whole genome sequencing of respiratory tract samples reveals full viral genomic information. J Clin Virol 2015;66:6–11.

[23] Nolan T, Borja-Tabora C, Lopez P, Weckx L, Ulloa-Gutierrez R, Lazcano-Ponce E, et al. Prevalence and incidence of respiratory syncytial virus and other respiratory viral infections in children 6 months to 10 years of age with influenza-like illness enrolled in a randomized trial. Clin Infect Dis 2015;60:e80–9.

[24] Wylie KM, Mihindukulasuriya KA, Sodergren E, Weinstock GM, Storch GA. Sequence analysis of the human virome in febrile and afebrile children. PLoS One 2012:e27735.

[25] Byington CL, Ampofo K, Stockmann C, Adler FR, Herbener A, Miller T, et al. Community surveillance of respiratory viruses among families in the Utah Better Identification of Germs-Longitudinal Viral Epidemiology (BIG-LoVE) Study. Clin Infect Dis 2015;61(8): 1217–24.

[26] Focosi D, Maggi F, Albani M, Macera L, Ricci V, Gragnani S, et al. Torquetenovirus viremia kinetics after autologous stem cell transplantation are predictable and may serve as a surrogate marker of functional immune reconstitution. J Clin Virol 2010;47:189–92.

[27] Madsen CD, Eugen-Olsen J, Kirk O, Parner J, Kaae Christensen J, Brasholt MS, et al. TTV viral load as a marker for immune reconstitution after initiation of HAART in HIV-infected patients. HIV Clinical Trials 2002;3:287–95.

[28] Young JC, Chehoud C, Bittinger K, Bailey A, Diamond JM, Cantu E, et al. Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. Am J Transplant 2015;15: 200–9.

[29] Phan TG, Kapusinszky B, Wang C, Rose RK, Lipton HL, Delwart EL. The fecal viral flora of wild rodents. PLoS Pathog 2011;7:e1002218.

[30] Taboada B, Espinoza MA, Isa P, Aponte FE, Arias-Ortiz MA, Monge-Martínez J, et al. Is there still room for novel viral pathogens in pediatric respiratory tract infections? Plos One 2014:e113570.