Received: 10 May 2017 | Revised: 19 September 2017 | Accepted: 12 October 2017
DOI: 10.1002/jmv.24984

RESEARCH ARTICLE

WILEY JOURNAL OF MEDICAL VIROLOGY

Metagenomic analysis of viral diversity in respiratory samples from patients with respiratory tract infections in Kuwait

Nada Madi[1] | Widad Al-Nakib[1] | Abu Salim Mustafa[1] | Nazima Habibi[2]

[1] Virology Unit, Department of Microbiology, Faculty of Medicine, Kuwait University, Safat, Kuwait

[2] Research Core Facility and OMICS Research Unit, Faculty of Medicine, Kuwait University, Safat, Kuwait

Correspondence
Dr Nada Madi, Assistant Professor, Department of Microbiology, Faculty of Medicine, Kuwait University, P.O.Box 24923, Safat 13110, Kuwait.
Email: madi@hsc.edu.kw

Funding information
Research Sector, Office of Vice President for Research, Kuwait University, Grant number: MI03/15

A metagenomic approach based on target independent next-generation sequencing has become a known method for the detection of both known and novel viruses in clinical samples. This study aimed to use the metagenomic sequencing approach to characterize the viral diversity in respiratory samples from patients with respiratory tract infections. We have investigated 86 respiratory samples received from various hospitals in Kuwait between 2015 and 2016 for the diagnosis of respiratory tract infections. A metagenomic approach using the next-generation sequencer to characterize viruses was used. According to the metagenomic analysis, an average of 145, 019 reads were identified, and 2% of these reads were of viral origin. Also, metagenomic analysis of the viral sequences revealed many known respiratory viruses, which were detected in 30.2% of the clinical samples. Also, sequences of non-respiratory viruses were detected in 14% of the clinical samples, while sequences of non-human viruses were detected in 55.8% of the clinical samples. The average genome coverage of the viruses was 12% with the highest genome coverage of 99.2% for respiratory syncytial virus, and the lowest was 1% for torque teno midi virus 2. Our results showed 47.7% agreement between multiplex Real-Time PCR and metagenomics sequencing in the detection of respiratory viruses in the clinical samples. Though there are some difficulties in using this method to clinical samples such as specimen quality, these observations are indicative of the promising utility of the metagenomic sequencing approach for the identification of respiratory viruses in patients with respiratory tract infections.

KEYWORDS
human metapneumovirus, influenza virus, research and analysis methods, respiratory syncytial virus, RNA extraction

1 | INTRODUCTION

Historically, the word metagenome was first used in 1998 to describe the collection of microbial genome found in a soil sample, including microorganisms that could not be cultured by conventional methods.[1] Early metagenomics studies on environmental samples yielded the identification of metabolic characters, the classification of organisms and the discovery of antibiotics and enzymes.[2,3] Now, metagenomic studies include a broad range of research including marine ecological research, plant and agricultural, human genetics and diagnostics of human diseases. The first application of metagenomic in virus discovery was in the analysis of viruses in soil samples from two marine sites in San Diego.[4] Later, the approach was used to analyze viruses in different fields.[5] The traditional methods of virus detection involved filtration, tissue culture, electron microscopy, and serology.[6] Among these, the standard gold method of virus detection was cell culture. However, many viruses cannot be easily cultivated, and two milestone innovations solved this problem; polymerase chain reaction (PCR) and DNA sequencing (Sanger method).[6] By using these methods, several important emerging viruses, such as hendra virus,[7] nipah virus,[8] menangle virus,[9,10] melaka virus,[11] and reston ebola virus,[12] were discovered. Despite the sensitivity of PCR, it can only detect one virus at a time. Multiplex PCR, on the other hand, is used to identify multiple targets for the identification of more than one virus in a single test. However, it is often difficult to standardize the assay by using various primers. Hence, novel approaches that overcome the difficulties of viral detection with the conventional molecular methods are needed to discover novel human viruses.[13] Not surprisingly, virologists were the first to explore the use of sequence-independent, a metagenomic approach using Next Generation Sequencer (NGS) to detect human-associated viruses. The genome of DNA and RNA viruses can be detected directly after extraction from samples through metagenomic sequencing.[14]

Worldwide, respiratory tract infection (RTI) is an important cause of hospitalization among young children and elderly, with significantly high mortality and morbidity.[15] RTI is a group of diseases of both upper or lower respiratory tract. Upper respiratory tract infections (URTIs) include laryngitis, common cold, rhinitis, pharyngitis/tonsillitis, otitis media and rhinosinusitis/sinusitis. On the other hand, lower respiratory tract infections (LRTIs) include bronchitis, bronchiolitis, tracheitis and pneumonia. These respiratory tract infections increase the extent of the problem in patients with chronic comorbidities and asthma,[16] chronic obstructive pulmonary disease (COPD),[17] very young, elderly,[17] and immunocompromized patients. Respiratory viruses account for over 65% of all respiratory infections and 90% of URTIs.[18,19] Bacteria, however, only represent 10% of all URTIs.[19] Despite the improvement in the molecular techniques for viral discovery, the viral aetiology of RTI is still unknown in a high number of cases either because the tests are ineffective or the virus is unrelated to any of the currently known respiratory viruses. Current diagnostic methods to detect respiratory viruses are mainly based on sequence-dependent molecular amplification techniques such as PCR, which identify a panel of known viruses, and therefore, new respiratory viruses are missed. To overcome the pitfalls associated with PCR approaches based on NGS techniques such as viral metagenomic will become a logical step as routine viral diagnostics on clinical samples. Broaden not only the detection range of viruses but also provide an additional characterization of the detected viruses such as genotypes and subtypes of viruses. However, the efficiency and viability of using such techniques in diagnostic setting require further study.

Since there are no studies on the use of metagenomic approach for the identification of pathogens responsible for different diseases, the novelty of this study was to develop a metagenomic approach using NGS for the detection of known and unknown viruses associated with respiratory tract infections among patients in Kuwait.

2 | METHODOLOGY

2.1 | Study population and sample collection

A total of 86 patients (56 were females and 30 males) with signs and symptoms of URTIs and LRTIs between 2015 and 2016 were enrolled in this study. The patients were admitted to different hospitals in Kuwait including Al-Sabah Hospital, Al-Farwaniya Hospital, Al-Adan Hospital, Mubarak Al-Kabeer Hospital, Infectious Diseases Hospital, and Al-Amiri Hospital. The respiratory diseases of the patients were pharyngitis/tonsillitis, rhinitis, bronchitis, bronchiolitis, and pneumonia. The patients ranged in age from one day to 89 years, with a median age of two years. Summary of clinical samples is shown in Table 1. The respiratory samples included nasopharyngeal aspirates/wash, nasopharyngeal swab, bronchoalveolar lavage, tracheal aspirates, sputum, throat swabs, and nasal swabs. They were collected in the hospital and processed at the Virology Unit, Faculty of Medicine, Kuwait University for the presence of viral nucleic acids using multiplex Real-Time PCR and metagenomic analysis based on next-generation sequencing.

2.2 | Nucleic acids extraction and multiplex real-time PCR assay

Total nucleic acids were isolated from clinical samples using Roche ®MagNA Pure LC system (Roche Diagnostics, Indianapolis, IN), according to the manufacturer's instructions and stored at −80°C for further processing. Multiplex Real-Time PCR assay using Fast Track Kit (Fast −Track Diagnostic, Luxembourg, Germany) was used to detect common respiratory viruses from the extracted RNA according to manufacturers' instructions. This assay is Multiplex Real-Time PCR for detection of 21 respiratory viruses by TaqMan® technology. It required five tube multiplex for detection of influenza A, influenza A (H1N1) swl, influenza B, human rhinovirus (HRV), human coronavirus NL63 (HCoV-NL63), HCoV-229E, HCoV-OC43, HCoV-HKU1, parainfluenza (PIV) 1, 2, 3, 4, human metapneumovirus (HMPV) A/B, respiratory syncytial virus (RSV) A/B, bocavirus, adenovirus (AdV), parechovirus, enterovirus, mycoplasma pneumonia, and internal control. This PCR assay is a routine diagnostic test for the detection of respiratory viruses which is performed at the Virology Unit, Faculty of Medicine, Kuwait University.

2.3 | Next generation sequencing and metagenomic analysis

Fresh nucleic acids (RNA and DNA) were extracted from each respiratory sample processed for metagenomic analysis using the Illumina MiSeq (San Diego, CA) platform for NGS according to standard procedures.[20] Briefly, after nucleic acid extraction step, genomic and host DNA was removed from each sample using Ambion DNA-free (Life Technologies) according to manufacturer's instructions to obtain metagenome RNA. Then, 10 ng of DNA- free RNA was integrated into first-strand cDNA synthesis primed by random hexamers and then amplified using whole transcriptome amplification kit (Qiagen,

TABLE 1 Summary of metagenomic sequencing and multiplex real-time PCR data

Sample no.	Age	Gender	Total read[a]	% of genome coverage[b]	Metagenomic results[c]	Multiplex real-time PCR results	C_T[d]
1	4 yr	M	94 071	-	-	Human bocavirus	12.28
2	6 mo	M	94 183	-	-	Rhinovirus	28.10
3	5 yr	M	135 071	-	-	Adenovirus	31.77
4	1 mo	M	181 304	-	-	Respiratory syncytial virus	19.40
5	4 mo	M	219 673	1.85	Influenza A virus (H9N2)	Rhinovirus	27.81
6	3 yr	M	119 744	-	-	Respiratory syncytial virus	18.65
7	6 mo	F	210 747	6.69	Torque teno virus 19	Enterovirus	24.33
8	1 yr	M	428 426	5.4	Respiratory syncytial virus	Respiratory syncytial virus	25.87
9	7 yr	M	184 918	2.69	Rotavirus F chicken	Adenovirus	18.22
10	1 yr	M	154 571	23.24	Human bocavirus	Adenovirus + human bocavirus	AdV: 14.5056 HBoV: 13.0494
11	4 yr	M	205 218	4.16	Respiratory syncytial virus	Respiratory syncytial virus	22.58
12	11 yr	M	209 072	-	-	Enterovirus	34.02
13	23 yr	M	126 583	-	-	Human metapneumo virus	32.63
14	1 mo	M	76 761	-	-	Influenza A+ human bocavirus	Flu A:25.0305 HBoV: 15.1864
15	4 yr	F	269 314	3.56	Respiratory syncytial virus	Respiratory syncytial virus	26.43
16	68 yr	M	192 990	-	-	Influenza B	16.51
17	2 yr	F	202 955	-	-	Respiratory syncytial virus	17.78
18	1 yr	M	136 219	1.47	Respiratory syncytial virus	Respiratory syncytial virus + Adenovirus	AdV: 17.3483 RSV: 15.2738
19	2 yr	M	259,599	-	-	Rhinovirus	11.09
20	58 yr	M	127 399	2.84	Rotavirus A	Influenza A virus	15.31
				1.92	Influenza A virus		-
21	1 yr	M	439 927	-	-	Parainflunza 1+ adenovirus	PIV1: 18.6867 AdV: 25.1402
22	3 yr	M	166 168	3.44	Respiratory syncytial virus	Respiratory syncytial virus	25.43 -
				1.86	Rotavirus strain E		
23	5 yr	F	158 839	-	-	Influenza A virus	33.27
24	57 yr	M	168 581	2.43	Adult diarrhoea rotavirus strainJ19	Respiratory syncytial virus	5.98
				1.4	Influenza A virus (H2N2)		
25	13 yr	F	40 774	-	-	Respiratory syncytial virus	29.78
26	3 yr	M	39 964	-	-	Human bocavirus	35.27
27	13 yr	M	164 746	-	-	Influenza A virus	17.32
28	32 yr	F	112 605	1.01	Torque teno midi virus 2	Influenza A virus	14.55
29	1 mo	M	85 792	-	-	Respiratory syncytial virus + parainfluenza 3 virus	RSV: 32.9743 PIV3: 30.0936
30	2 yr	M	164 227	-	-	Respiratory syncytial virus	19.8748
31	4 mo	M	40 948	-	-	Rhinovirus	11.5602
32	4 yr	F	4 074	-	-	Influenza A virus	14.1571
33	1 mo	F	159 476	-	-	Respiratory syncytial virus	15.0062
34	3 yr	F	40 894	-	-	Respiratory syncytial virus	12.5713
35	1 mo	F	175 935	3.01	Rotavirus B	Respiratory syncytial virus	12.6546
				1.95	Influenza C virus		
36	4 mo	M	40 795	-	-	Respiratory syncytial virus	12.3435
37	1 mo	M	119 447	44.43	Human rhinovirus 14	Rhinovirus + human metapneumo virus	HRV: 11.5732 HMPV: 14.87
38	67 yr	M	40 993	1.76	Human respiratory syncytial virus	Influenza B + human respiratory syncytial virus	Flu B: 12.5212 RSV: 12.1386

(Continues)

TABLE 1 (Continued)

Sample no.	Age	Gender	Total read[a]	% of genome coverage[b]	Metagenomic results[c]	Multiplex real-time PCR results	C_T[d]
39	59 yr	M	209 709	-	-	Influenza A virus	15.6758
40	89 yr	M	117 703	-	-	Influenza A virus	15.3729
41	25 yr	F	107 994	2.06	Human respiratory syncytial virus	Respiratory syncytial virus	12.5934
42	2 yr	M	257 959	99.23	Respiratory syncytial virus	Respiratory syncytial virus	10.36
43	70 yr	F	100 377	-	-	Adenovirus	12.37
44	1 mo	M	173 720	2.03	Influenza C virus	Rhinovirus	11.96
45	4 mo	M	33 958	-	-	Respiratory syncytial virus	21.91
46	66 yr	F	22 444	3.01	Human rotavirus B strain	Influenza A virus	13.48
47	1 mo	M	301 529	-	-	Respiratory syncytial virus	19.0166
48	2 mo	M	187 769	85.88	Bovine corona virus	Coronavirus OC 43	16.1313
				72.52	Human metapneumo virus		
				23.73	Coronavirus HKU14		
49	2 yr	F	24 825	-	-	Human metapneumo virus	24.3051
50	47 yr	M	24 726	-	-	Human metapneumo virus	11.8083
51	1 yr	M	168 915	-	-	Rhinovirus	33.9199
52	6 yr	F	18 995	-	-	Resipiratory syncytial virus	12.574
53	28 yr	F	100 404	-	-	Human metapneumo virus	12.4938
54	16 yr	M	121 120	2.57	Human rhinovirus C	Rhinovirus	11.6254
55	68 yr	F	116 653	4.05	Respiratory syncytial virus	Influenza A virus	18.5788
56	2 mo	M	221 941	-	-	Rhinovirus	13.0581
57	1 yr	M	74 807	1.97	Human respiratory syncytial virus	Adenovirus	12.466
58	1 mo	M	167 451	3.33	Respiratory syncytial virus	Respiratory syncytial virus	12.525
59	11 mo	M	88 049	-	-	Human metapneumo virus	12.1959
60	63 yr	F	60 882	-	-	Rhinovirus	21.5916
61	27 yr	F	141 599	5.2	Rotavirus C	Rhinovirus	28.7566
62	9 yr	M	177 528	-	-	Rhinovirus	11.6255
63	10 yr	F	164 350	77.21	Respiratory syncytial virus	Respiratory syncytial virus	17.1614
64	1 mo	F	159 543	2.47	Human rhinovirus C	Rhinovirus	22.5983
65	2 mo	F	306 199	-	-	-	-
66	4 yr	F	156 745	1.37	Human respiratory syncytial virus	-	-
67	14 dy	F	5 806	-	-		-
68	2 mo	M	302 314	4	Rotavirus strain E	-	-
69	4 mo	M	327 618	-	-	-	-
70	15 dy	F	137 021	2.63	Respiratory syncytial virus	-	-
71	5 mo	M	258 050	2.13	Influenza A virus (H1N1)	-	-
72	73 yr	M	323 340	2.82	Rotavirus D chicken	-	-
				1.85	Influenza A virus (H5N1)		-
73	1 yr	M	32 382	-	-	-	-
74	1 yr	F	18 984	-	-	-	-
75	2 yr	F	142 437	-	-	-	-
76	2 mo	F	39 441	-	-	-	-
77	1 mo	M	232 798	2.94	Rotavirus G chicken	RSV	13.45
				2.35	Hantaan virus		
78	2 mo	M	187 421	1.99	Rotavirus strain E	-	-
79	9 yr	M	43 808	-	-	-	-

(Continues)

TABLE 1 (Continued)

Sample no.	Age	Gender	Total read[a]	% of genome coverage[b]	Metagenomic results[c]	Multiplex real-time PCR results	C_T[d]
80	49 yr	M	35 596	-	-	-	-
81	12 yr	M	267 407	-	-	-	-
82	4 yr	M	114 665	2.33	Rotavirus strain E	-	-
83	3 dy	F	176 085	-	-	-	-
84	7 mo	M	180 162	2.92	Rotavirus strain E	-	-
85	7 mo	F	3	-	-	-	-
86	65 yr	M	45 363	-	-	-	-

[a]Total number of MiSeq reads after removal ow quality, short or adapter containing reads.
[b]All read mapped to deno novo assembled genome using Burrows-Wheel Aligner and the value reports the percentage of times each position sequenced.
[c]As per Burrows-Wheel Aligner results.
[d]C_T Real-Time PCR cycle threshold value.

Valencia, CA). An exact concentration of 1 μg of DNA was prepared after DNA quantification step using Qubit® 2.0 Fluorometer and Qubit™ dsDNA BR Assay kits (Invitrogen, CA). Illumina TruSeq DNA library preparation kit V2 (Illumina San Diego, CA) was used to prepare DNA libraries, followed by sequencing of 150-bp paired-end reads on an Illumina MiSeq instrument at the OMICS Research Unit, Health Science Centre, Kuwait University.

2.4 | Bioinformatics

The quality of the sequence data from a total of 86 Miseq runs was checked using FastQC (version 0.10.1; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The Sequence fastq files were trimmed where the average quality score was <30 using a fast toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html), and from bases 18-145 were kept for all files. High-quality paired-end sequences were retained for downstream analysis. Host and bacterial sequences were removed from these sequences by mapping them to a database containing all human and bacterial DNA/RNA sequences. A database of references containing about 5800 strains available on National Center for Biotechnology Information (NCBI) was created by merging all sequence into one FASTA file. The host removed reads were aligned against a viral reference using Burrows-Wheeler Aligner (BWA) and sam files were created. Option samtools flagstat provided the percentage of host and viral reads in sam files. A statistic summary providing the information regarding some reads corresponding to the individual viral strain and the percentage of viral genome covered for all viruses was extracted from sam files using pileup option in bb map program (http://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmap-guide/).

2.5 | Statistical analysis

The data were analyzed using a computer software "Statistical Package for Social Sciences," SPSS version 24.0 (IBM Corp, Armonk, NY). The descriptive statistics are presented as frequencies and percentages. The Cohen's Kappa (k) was applied to find the measure of agreement on detecting respiratory viruses by both multiplex Real-Time PCR and metagenomics sequencing techniques. The two-tailed probability value "p" <0.05 was considered statistically significant.

3 | RESULTS

3.1 | Virus detection by multiplex real-time PCR

Eighty-six respiratory samples from patients with URTIs and LRTIs who visited different hospitals in Kuwait between 2015 and 2016 were analyzed for respiratory viruses by multiplex Real-Time PCR assay. The results showed that 65 out of 86 samples (75.6%) were positive for respiratory viruses. Among the positive samples, 23 (35.4%) were positive for RSV, 12 (18.5%) samples were positive for HRV, eight (12.3%) samples were positive for influenza A, four (6.2%) samples were positive for AdV, four (6.2%) samples were positive for HMPV, two (3%) samples were positive for bocavirus, two (3%) samples were positive for enterovirus, one (1.5%) sample was positive for influenza B, one (1.5%) sample was positive for PIV, one (1.5%) sample was positive for HCoV-OC43, and seven (10.7%) samples were positive for dual viruses (Figure 1). The combinations of dual virus samples were as follow: adenovirus and human bocavirus; human bocavirus and influenza A virus; RSV and AdV; AdV and PIV-A virus; RSV and PIV-3; HRV and HMPV; RSV and influenza B virus. The remaining 21 (24.4%) samples were negative for any respiratory viruses (Figure 1). The threshold cycle value (C_T) for each detected virus is shown in Table 1.

3.2 | Virus detection by metagenomic sequencing

From the 86 samples, an adequate amount of RNA was obtained to perform whole-transcriptome sequencing using the Illumina Miseq sequencer for the detection of RNA viruses. Sequencing of cDNA libraries from all samples generated an average of 145, 019 reads (range; 3-439 927), after quality trimming and filtering. On an average, 64% (range, 5.72-98.16%) of the reads were derived from host (human) genome, and 2% (range, 0.04-7.87%) of the reads were derived from viruses (Figure 2). The rest of the reads (34%) were derived from bacteria and other viruses that did not have any

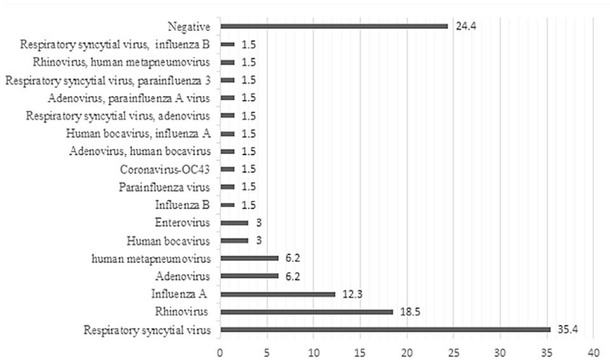

FIGURE 1 Clustered Bar chart of the percentage of patients positive for respiratory viruses detected by multiplex RT-PCR assay ($n = 86$)

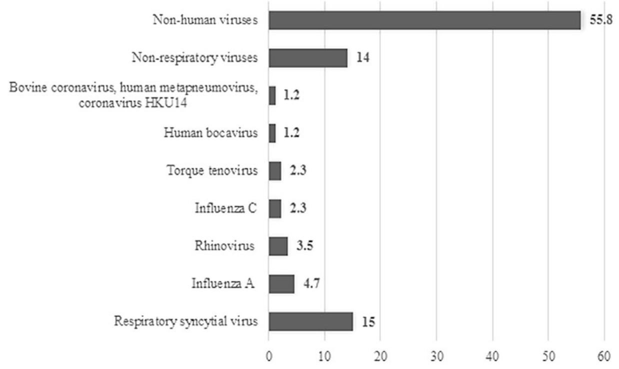

FIGURE 3 Clustered Bar chart of the percentage of patients positive for viruses detected by metagenomic sequencing assay ($n = 86$)

significant match in the known databases. The identified viral sequences belonged to the following respiratory viruses; respiratory syncytial virus (15% of patient), influenza A virus (4.7% of patient), rhinovirus (3.5% of patient), influenza C virus (2.3% of patient), torque tenoviruses(2.3% of patient), human bocavirus (1.2% fo patient), mixed viral infection consisted of bovine coronavirus, human metapneumovirus, and coronavirus HKU14 (1.2% of patient) (Figure 3). Also, non-respiratory viruses were detected in 14% of the patients. These viruses were rotaviruses A, B, C, D, E, F, G, J19, and hantavirus. The rest of the identified viral sequences belonged to plenty of non-human viruses, such as plant associated viruses, were found in 55.8% of the patients (Figure 3). The average genome coverage of the identified viruses after genome sequence assembly was 12%. The highest genome coverage was for RSV (99.23%), and the lowest for torque teno midi virus 2 (1.01%) (Table 1). Table 1 demonstrates a summary of the metagenomic sequencing and multiplex Real-Time PCR results for each clinical sample. It is worth mentioning that sequences from human endogenous retrovirus, which is a non-pathogenic virus, was common shared viral sequences found in 44 sample (51.2%) with an average genome coverage of 5.2% (range, 1.2-11.6%).

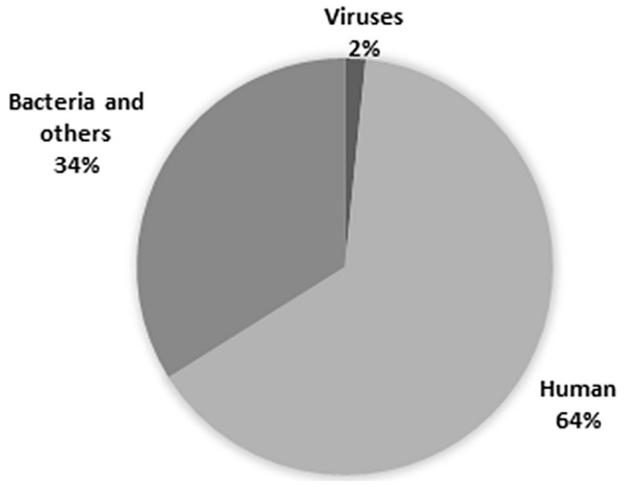

FIGURE 2 Pie chart of the taxonomic distribution of metagenomic sequencing reads from clinical samples. Data are average values of reads

To evaluate the feasibility of the metagenomic approach for the detection of respiratory viruses in clinical samples, a comparison was made between multiplex Real-Time PCR and metagenomic sequencing results. Of the total 86 clinical specimens, multiplex Real-Time PCR detected 65 (75.6%) clinical samples as positive for respiratory viruses and 21 (24.4%) as negative, while metagenomic sequencing technique detected 28 (32.6%) clinical samples positive and 58 (67.4%) negative. Overall, 24 (27.9%) clinical samples were detected as positive, and 17 (19.8%%) as negative, by both the techniques, giving an absolute agreement on 41 (47.7%) clinical samples. However, applying the Cohen's Kappa statistics for a measure of agreement ($k = 0.112$, $P = 0.129$), only a slight agreement was found between the two techniques.

4 | DISCUSSION

The respiratory tract is a target of many human viruses, especially RNA viruses. The current molecular techniques for the detection of respiratory viruses are largely targeted dependent tests which detect a limited number of viruses. On the other hand, NGS-based metagenome approaches are target independent which can detect common, unexpected pathogens, and novel viruses in a given sample.[21]

To explore the utility of the metagenomics approach in the identification of respiratory viruses in clinical samples, we first performed multiplex Real-Time PCR testing for the initial identification of respiratory viruses in respiratory samples from patients with URTIs and LRTIs. RSV was the most predominant respiratory virus detected in the patient's samples (35.4% of the patients), while HRV was the second prevalent virus (18.5% of the patients). Our observation is in agreement with others who showed that RSV is the most prevalent virus in patients with respiratory tract infections.[22–24] In addition to single viral infections, our result demonstrated that mixed viral infections are a frequent phenomenon; it was observed in seven samples indicating that more than one virus can trigger respiratory tract infections and that may influence the severity and outcomes of the respiratory disease. One patient with mixed viral infection presented with severe bronchopneumonia while the rest of the patients had mild respiratory symptoms. These results suggested the absence of an association between multiple viral infections and the severity of the respiratory disease. In contrast to our results, studies showed that patients with respiratory tract infection due to dual viral infections were hospitalized significantly more often than those with respiratory disease due to a single virus.[25,26] In agreement with our results, others did not demonstrate any correlation between the presence of multiple viruses and the severity of the respiratory diseases.[27–30]

Respiratory samples from the 86 patients were analyzed further for the presence of viruses by metagenomic sequencing using NGS. It should be noted that this study is the first of its kind to perform metagenomic analysis of the virome of respiratory tract specimens. Sequencing of cDNA libraries from the clinical samples revealed an average of 145 019 of 150-bp paired-end reads which considered relatively lower than expected. We anticipated that this problem is due to the low concentration of the cDNA obtained by the whole transcriptome amplification step. Analysis of the sequencing reads revealed that over half of the reads (64%) were host-derived reads, however, only 2% of the reads derived from viruses. The residual of the reads (34%) included bacterial reads and other reads that did not have any significant match in the database. Other researchers have also reported high percentage (90%) of human-derived sequence in nasal specimens analyzed by metagenomics sequencing.[31–33] Therefore, it appears that the presence of human-derived reads is an intrinsic problem when nucleic acids are extracted directly from respiratory samples. In comparison to our results, Nakamura and coworkers identified lower average (0.76%) of virus-associated sequences, and Yang and colleagues identified only 0.05% of virus-associated sequences.[32,34]

Corresponding with multiplex Real-Time PCR assay, RSV was the most predominant virus detected in the respiratory samples by the NGS-based metagenomic approach, while influenza A virus was the second most prevalent virus. Interestingly, two torque teno viruses were detected by the metagenomic approach in two respiratory samples; torque teno virus 19 and torque teno midi virus 2. Torque teno viruses (TTV) are new, emerging infectious agents that are exceptionally high prevalence, and relatively uniform distribution worldwide.[35,36] It was suggested that TTV is related to numerous diseases, such as hepatitis, respiratory diseases, cancer, haematological and autoimmune disorders. However, their direct involvement is under arguments.[37–39] Metagenomic sequencing in this study revealed mixed viral infection (bovine coronavirus, HMPV, and HCoV-HKU14) in only one clinical sample.

Many non-respiratory viruses were detected in 14% of the clinical samples. Different strains of rotaviruses were among these viruses. Indeed, it is not unusual to detect rotaviruses in respiratory specimens. It was shown in the earlier studies the detection of rotavirus in nasopharyngeal secretions of children with acute gastritis. These studies speculated the association of diarrhoea due to rotavirus with respiratory signs and symptoms.[40,41] Another prominent observation in our study was the detection of a vast number of non-human viruses such plant viruses in all clinical samples. We speculate that the presence of plant viruses in the respiratory samples is a natural phenomenon. They come from food, water, or dust and eventually become be part of normal flora of the upper respiratory tract which may/or may not be linked to any disease. Another prominent observation was the detection of sequences from a human endogenous retrovirus (HERV) approximately half of the patients. This observation reveals the high prevalence of human endogenous retrovirus in the population of Kuwait. According to many investigations, it was found that human endogenous viruses are old viruses that started to integrate into the human genome around 30-40 million years ago and now about 8% of the human genome is derived from sequences with similarity to retroviruses.[42]

One of the major questions is how to intercept metagenomic analysis results by NGS in term what is clinically relevant for the patient. The comparison of the metagenomic sequencing data to RT-Real time PCR data revealed absolute agreement of 47.7%. Although metagenomic approach failed to detect all the multiplex PCR positive samples, it did offer several advantages over the multiplex Real-Time PCR method. For examples, the metagenomic sequencing identified several viruses causing respiratory infections that are usually not detected using routine diagnostic assays (eg, torqueteno viruses and coronavirus HKU14). Also, it delivered more detailed typing information for the detected viruses including subtype for some viruses. Furthermore, it is possible that multiplex RT-PCR amplifies viruses present in minimal concentrations that have no significance in the disease process. It should be noted that two nucleic acid extracts for each clinical sample were prepared. The first extract was used to perform RT-PCR and the second extract was used to perform metagenomics sequencing using NGS. This, in turn, will generate differences in nucleic acid concentration and quality which may further affect the results.

The analytical sensitivity of metagenomics sequencing is highly influenced by the quality of the samples. To improve the sensitivity of this approach is by increasing the depth of sequencing, which in turn can be overcome by improving the sample preparation.[43] One of the solutions is to perform additional steps of ample-filtering before sequencing to eliminate host cells and enhance the microbial genome to ensure more efficient microbial pathogen detection. Some studies employed a pre-treatment step before nucleic acid extraction which required a combination of filtration and density-dependent centrifugation to enhance the majority of viruses.[4,44,45] Others used an enzyme cocktail containing DNasI, RNaseI, and benzonase to digest naked host nucleic acid.[46] One of the limitations of pre-treatment step to remove host DNA contents from clinical samples is the possibility to introduce a bias and decrease the sensitivity for DNA viruses such as adenovirus which considered as vital virus causing respiratory diseases. Also, filtering steps may reduce significantly the quality of DNA/RNA produced, which may further effect metagenomics sequencing results. Therefore, it is recommended to keep pre-treatment steps to a minimum to avoid any particular bias. We have used a virus identification algorithm, BWA and BLAST from NCBI for the identification, clustering and read removal. One noticeable step for future improvement lies in the treatment of the BLAST undefined reads. BLAST search did not classify a large portion of the sequence data generated here. Future efforts will include an emphasis on alternative virus identification to characterize these novel sequences.

Massive sequencing projects and databases based on NGS technologies are not safe from contamination issues.[47,48] It is problematic when contaminant sequence reads being lost in the numerous of reads from the target sample, and consequently difficult to detect and clean out. In NGS library assembly protocols, it is essential to have one or multiple PCR amplification steps that could elevate the concentrations of DNA, thereby increasing the risk of contamination. Also, cross-contamination of starting material is a devastating issue. There is a chance that samples may contaminate each other when they are prepared at the same time, or multiple libraries are produced in parallel. To avoid this type of contamination is to perform all sample and library preparations individually. If it is not feasible to do all sample and library preps separately, then strict measures are needed to be employed to keep samples pure. Certainly, it is essential to identify at which steps of the experimental protocol the contaminations are most often happening, and to deliver guidelines on how to avoid it.

5 | CONCLUSION

In conclusion, we conducted a metagenomic sequencing analysis of the viruses in patients with RTI. The results presented in this study demonstrated that metagenomic approach is a promising diagnostic tool in clinical virology in which RT-Real-Time PCR couldn't detect many respiratory viruses that were detected by metagenomic approaches. Such sequence independent detection method will increase the chance to detect the causative agent of viral RTI and to gain information of the viruses that cannot be obtained by the current diagnostic tool. Although many obstacles to the routine use of metagenomic approaches do exist, which include cost, labor intensity and turnover time, it was proved to be highly sensitive in many studies, and it theoretically provides more information regarding virus species/type for virus diagnosis with the ability of the detection of unknown viruses. The question is probably when metagenomics sequencing will become a routine test for detecting infectious viruses? This approach requires improvement in sample preparation, validated pipelines for read sorting and taxonomic assignation and soon will substitute the current diagnostic tools. Indeed, this novel study provided outstanding information on the association of different viruses with RTIs in Kuwait. Moreover, metagenomics queening approach will be used in the near future to study the viral diversity in other syndromes such as pyrexia of unknown origin and gastroenteritis.

ACKNOWLEDGMENT

Research Sector, Kuwait University, funded this project (Project No. MI03/15). We extend our thanks to the Research Core Facility and OMICS Research Unit for their tremendous support and help in using the facilities provided in the Unit. Special thanks to Ms Anfal Al-Adwani for here technical assistance. We are also grateful to Mr Mohammed Khan for his outstanding help in the bioinformatics analysis of the data. Special thanks to Dr Prem Sharma (Technical Advisor, Vice President Office, Kuwait University) for his help in the statistical analysis of the data.

ORCID

Nada Madi http://orcid.org/0000-0002-4654-544X

REFERENCES

1. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol*. 1998;5: R245–R249.
2. Riesenfeld CS, Schloss PD, Handelsman J. Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet*. 2004;38:525–552.
3. Krause DO, Denman SE, Mackie RI, et al. Opportunities to improve fiber degradation in the rumen: microbiology, ecology, and genomics. *FEMS Microbiol Rev*. 2003;27:663–693.
4. Breitbart M, Salamon P, Andresen B, et al. Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci USA*. 2002;99:14250–14255.
5. Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. *Curr Opin Virol*. 2012;2:63–77.
6. Xie G, Yu J, Duan Z. New strategy for virus discovery: viruses identified in human feces in the last decade. *Sci China Life Sci*. 2013;56:688–696.
7. Murray K, Rogers R, Selvey L, et al. A novel morbillivirus pneumonia of horses and its transmission to humans. *Emerg Infect Dis*. 1995;1:31–33.
8. Chua KB, Bellini WJ, Rota PA, et al. Nipah virus: a recently emergent deadly paramyxovirus. *Science*. 2000;288:1432–1435.
9. Philbey AW, Kirkland PD, Ross AD, et al. An apparently new virus (family Paramyxoviridae) infectious for pigs, humans, and fruit bats. *Emerg Infect Dis*. 1998;4:269–271.
10. Chua KB, Crameri G, Hyatt A, et al. A previously unknown reovirus of bat origin is associated with an acute respiratory disease in humans. *Proc Natl Acad Sci USA*. 2007;104:11424–11429.
11. Chua KB, Crameri G, Hyatt AD, et al. A previously unknown reovirus of bat origin is associated with an acute respiratory disease in humans. *Proc Natl Acad Sci*. 2007;104:11424–11429.
12. Barrette RW, Metwally SA, Rowland JM, et al. Discovery of Swine as a Host for the Reston ebolavirus. *Source Sci New Ser*. 2009;325:204–206.
13. Denno DM, Klein EJ, Young VB, Fox JG, Wang D, Tarr PI. Explaining unexplained diarrhea and associating risks and infections. *Anim Health Res Rev*. 2007;8:69–80.
14. Pallen MJ. Diagnostic metagenomics: potential applications to bacterial, viral and parasitic infections. *Parasitology*. 2014;141:1–7.
15. Armstrong GL, Conn LA, Pinner RW. Trends in infectious disease mortality in the United States during the 20th century. *JAMA*. 1999;281:61–66.
16. Nicholson KG, Kent J, Hammersley V, Cancio E. Acute viral infections of upper respiratory tract in elderly people living in the community: comparative, prospective, population based study of disease burden. *BMJ*. 1997;315:1060–1064.
17. Gorse GJ, O'Connor TZ, Hall SL, Vitale JN, Nichol KL. Human coronavirus and acute respiratory illness in older adults with chronic obstructive pulmonary disease. *J Infect Dis*. 2009;199:847–857.

18. Lodes MJ, Suciu D, Wilmoth JL, et al. Identification of upper respiratory tract pathogens using electrochemical detection on an oligonucleotide microarray. *PLoS ONE*. 2007;2:e924.

19. Fahey T, Stocks N, Thomas T. Systematic review of the treatment of upper respiratory tract infection. *Arch Dis Child*. 1998;79:225–230.

20. Moore NE, Wang J, Hewitt J, et al. Metagenomic analysis of viruses in feces from unsolved outbreaks of gastroenteritis in humans. *J Clin Microbiol*. 2015;53:15–21.

21. Prachayangprecha S, Schapendonk CME, Koopmans MP, et al. Exploring the potential of next-generation sequencing in detection of respiratory Viruses. *J Clin Microbiol*. 2014;52:3722–3730.

22. Gamiňo-Aryo AE, Moreno-Espinosa S, Llamosas-Gallardo B, et al. Epidemiology and clinical characteristics of respiratory syncytial virus infections among children and adults in Mexico. *Influenza Other Respi Viruses* 2017;11:48–56.

23. Gurgel R, Bezerra P, Duarte Mdo C, et al. Relative frequency, possible risk factors, viral codetection rates, and seasonality of respiratory syncytial virus among children with lower respiratory tract infection in northeastern Brazil. *Medicine*. 2016;95:1–8.

24. El Kholy AA, Mostafa NA, Ali AA, et al. The use of multiplex PCR for the diagnosis of viral severe acute respiratory infection in children: a high rate of co -detection during the winter season. *Eur J Clin Microbiol Infect Dis*. 2016;35:1607–1613.

25. Semple MG, Cowell A, Dove W, et al. Dual infection of infants by human metapneumovirus and human respiratory syncytial virus is strongly associated with severe bronchiolitis. *J Infect Dis*. 2005; 191:382–386.

26. Drews a L, Atmar RL, Glezen WP, Baxter BD, Piedra PA, Greenberg SB. Dual respiratory virus infections. *Clin Infect Dis*. 1997;25:1421142–1421149.

27. Asner SA, Science ME, Tran D, Smieja M, Merglen A, Mertz D. Clinical disease severity of respiratory viral co-infection versus single viral infection: a systematic review and meta-analysis. *PLoS ONE*. 2014;9: e99392.

28. Scotta MC, Chakr VCBG, de Moura A, et al. Respiratory viral coinfection and disease severity in children: a systematic review and meta-analysis. *J Clin Virol*. 2016;80:45–56.

29. Martin ET, Kuypers J, Wald A, Englund JA. Multiple versus single virus respiratory infections: viral load and clinical disease severity in hospitalized children. *Influenza Other Respi Viruses*. 2012;6:71–77.

30. De Paulis M, Gilio AE, Ferraro AA, et al. Severity of viral coinfection in hospitalized infants with respiratory syncytial virus infection. *J Pediatr (RioJ)*. 2011;87:307–313.

31. Greninger AL, Chen EC, Sittler T, et al. A metagenomic analysis of pandemic influenza a (2009 H1N1) infection in patients from North America. *PLoS ONE*. 2010;5.

32. Nakamura S, Yang CS, Sakon N, et al. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS ONE*. 2009;4.

33. Greninger AL, Chen EC, Sittler T, et al. A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PLoS ONE*. 2009;5:e13381.

34. Yang J, Yang F, Ren L, et al. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. *J Clin Microbiol*. 2011;49:3463–3469.

35. Gallian P, Berland Y, Olmer M, et al. TT virus infection in French hemodialysis patients: study of prevalence and risk factors. *J Clin Microbiol*. 1999;37:2538–2542.

36. Hsu HY, Ni YH, Chen HL, Kao JH, Chang MH. TT virus infection in healthy children, children after blood transfusion, and children with non-A to E hepatitis or other liver diseases in Taiwan. *J Med Virol*. 2003;69:66–71.

37. Spandole S, Cimponeriu D, Berca LM, Mihuaescu G. Human anelloviruses: an update of molecular, epidemiological and clinical aspects. *Arch Virol*. 2015;160:893–908.

38. Hino S, Miyata H. Torque teno virus (TTV): current status. *Rev Med Virol*. 2007;17:45–57.

39. Maggi F, Pifferi M, Tempestini E, et al. TT virus loads and lymphocyte subpopulations in children with acute respiratory diseases. *J Virol*. 2003;77:9081–9083.

40. Echeverria P, Blacklow NR, Cukor GG, Vibulbandhitkit S, Changchawalit S, Boonthai P. Rotavirus as a cause of severe gastroenteritis in adults. *J Clin Microbiol*. 1983;18:663–667.

41. Zheng BJ, Chang RX, Ma GZ, et al. Rotavirus infection of the oropharynx and respiratory tract in young children. *J Med Virol*. 1991;34:29–37.

42. Tugnet N, Rylance P, Roden D, Trela M, Nelson P. Human endogenous retroviruses (HERVs) and autoimmune rheumatic disease: is there a link?. *Open Rheumatol J*. 2013;7:13–21.

43. Eloit M, Lecuit M. The diagnosis of infectious diseases by whole genome next-generation sequencing: a new era is opening. *Front Cell Infect Microbiol*. 2014;4:25.

44. Angly FE, Felts B, Breitbart M, et al. The marine viromes of four oceanic regions. *PLoS Biol*. 2006;4:2121–2131.

45. Dinsdale EA, Edwards RA, Hall D, et al. Functional metagenomic profiling of nine biomes. *Nature*. 2008;452:629–632.

46. Kohl C, Brinkmann A, Dabrowski PW, Radonic A, Nitsche A, Kurth A. Protocol for metagenomic virus detection in clinical specimens. *Emerg Infect Dis*. 2015;21:48–57.

47. Lusk RW. Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. *PLoS ONE*. 2014;9:e110808.

48. Merchant S, Wood DE, Salzberg SL. Unexpected cross-species contamination in genome sequencing projects. *Peer J*. 2014; 2:e675.

How to cite this article: Madi N, Al-Nakib W, Mustafa AS, Habibi N. Metagenomic analysis of viral diversity in respiratory samples from patients with respiratory tract infections in Kuwait. *J Med Virol*. 2018;90:412–420. https://doi.org/10.1002/jmv.24984