



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

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

information on hepatitis C and other bloodborne viruses is also needed in non-healthcare settings, particularly in practices where sharp instruments such as slicers that are likely to cause hand injuries are used and shared by workers. In these practices, healthcare authorities should highlight the importance of wearing gloves and the prevention of cutting injuries.

## Acknowledgements

We are indebted to Dorothée Obach for her assistance in the preparation of the paper.

## Transparency Declaration

Y. Yazdanpanah received travel grants, honoraria for presentations at workshops and consultancy honoraria from Bristol-Myers Squibb, Boehringer Ingelheim, Gilead, Glaxo-SmithKline, Merck, Pfizer, Roche and Tibotec. All the other authors state the absence of dual/conflicting interests.

## References

1. Alberti A, Chemello L, Benvegna L. Natural history of hepatitis C. *J Hepatol* 1999; 31 (suppl 1): 17–24.
2. Alter MJ. Epidemiology of hepatitis C virus infection. *World J Gastroenterol* 2007; 13: 2436–2441.
3. Simmonds P, Bukh J, Combet C et al. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* 2005; 42: 962–973.
4. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; 22: 4673–4680.
5. Felsenstein J dbta. PHYLIP: phylogeny inference package, version 3.5c (computer program). Seattle, WA: Department of Genetics, University of Washington, 1993.
6. Perriere G, Gouy M. WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie* 1996; 78: 364–369.
7. Simmonds P. Genetic diversity and evolution of hepatitis C virus—15 years on. *J Gen Virol* 2004; 85: 3173–3188.
8. Alfonso V, Flichman DM, Sookoian S, Mbayed VA, Campos RH. Evolutionary study of HVRI of E2 in chronic hepatitis C virus infection. *J Gen Virol* 2004; 85: 39–46.
9. Anonymous. L'épidémiologie des hépatites B et C en France. *Bulletin Epidémiologique Hebdomadaire thématique* 2009; 20–21: 195.
10. Santantonio T, Medda E, Ferrari C et al. Risk factors and outcome among a large patient cohort with community-acquired acute hepatitis C in Italy. *Clin Infect Dis* 2006; 43: 1154–1159.
11. Brouard C, Delarocque-Astagneau E, Meffre C et al. Trends of hepatitis C screening in France through Rena-VHC and hepatology reference centres surveillance systems. *Bulletin Epidémiologique Hebdomadaire thématique* 2009; 20–21: 199.
12. Puro V, Shouval D. Hepatitis B, hepatitis C, and other blood-borne infections in healthcare workers—a VHPB symposium report—Rome, Italy, March 17–18, 2005. *J Viral Hepat* 2005; 14: 1–16.
13. Yazdanpanah Y, De Carli G, Miguères B et al. Risk factors for hepatitis C virus transmission to Health Care Workers after occupational exposure: a European case-control study. *Rev Epidemiol Sante Publique* 2006; 54 suppl 1: IS23–IS31.
14. Jagger J, Puro V, De Carli G. Occupational transmission of hepatitis C virus. *JAMA* 2002; 288: 1469.
15. Leiss JK, Ratcliffe JM, Lyden JT et al. Blood exposure among paramedics: incidence rates from the national study to prevent blood exposure in paramedics. *Ann Epidemiol* 2006; 16: 720–725.

## Direct pathogen detection from swab samples using a new high-throughput sequencing technology

H. Yongfeng\*, Y. Fan\*, D. Jie\*, Y. Jian, Z. Ting, S. Lilian and Q. Jin

State Key Laboratory for Molecular Virology and Genetic Engineering, Institute of Pathogen Biology, Chinese Academy of Medical Sciences, Beijing, China

## Abstract

The detection of emerging infectious diseases has been a continuing concern, especially with the novel influenza A (H1N1) viral pandemic of 2009. In the present study, we validated a 'second-generation' parallel sequencing platform for viral detection in swab samples collected during recent influenza virus infections in Beijing. This operation yielded millions of valid reads per sample and resulted in an almost complete spectrum of nucleotide information. Importantly, novel A (H1N1) and seasonal A (H3N2) influenza virus-derived sequences were detected without prior knowledge or use of genetic information in advance, suggesting that this approach could be a valuable tool for diagnosing emerging infectious diseases.

**Keywords:** Emerging infectious diseases, high-throughput sequencing technology, novel influenza A (H1N1) virus, Solexa system, viral detection

**Original Submission:** 1 March 2010; **Revised Submission:** 5 April 2010; **Accepted:** 6 April 2010

Editor: D. Raoult

**Article published online:** 15 April 2010

*Clin Microbiol Infect* 2011; **17:** 241–244

10.1111/j.1469-0691.2010.03246.x

**Corresponding authors:** Q. Jin, State Key Laboratory for Molecular Virology and Genetic Engineering, Institute of Pathogen Biology, Chinese Academy of Medical Sciences, No. 6 Rongji Dongjie, BDA, Beijing 100176, China  
**E-mail:** zdsys@vip.sina.com

\*These authors contributed equally to this work.

The detection of emerging infectious diseases, such as severe acute respiratory syndrome, highly pathogenic avian influenza H5N1 and novel influenza A (H1N1), has become a continuous public health concern. Thus, the accurate diagnosis of pathogens in samples is becoming increasingly important. Currently, nucleic acid amplification tests (NAATs), such as PCR, nucleic acid sequence-based amplification, loop-mediated isothermal amplification and DNA microarray, are used with greater frequency in preference to traditional culture- or antigen-based diagnostic procedures, primarily as a result of their greater sensitivity. However, almost all diagnostic NAATs require viral genome information in advance, and thus cannot be used to detect or characterize novel or unexpected viral infections. Likewise, these methods are completely ineffective if a viral genome has evolved sufficiently to result in point mutations at primer binding sites [1,2].

New sequencing technologies referred to as 'second-generation,' such as 454 (Roche, Mannheim, Germany), Solexa (Illumina, Inc., San Diego, CA, USA) and SOLiD (Applied Biosystems, Foster City, CA, USA), show promise for the unbiased detection of pathogens. These systems allow researchers to obtain millions of sequences in a single round of operation, as well as an almost complete spectrum of nucleotide information through comparison and analysis with relevant databases. Because the 454 system has previously identified viral pathogens in nasal and faecal specimens [3], we tested the ability of Solexa, another unbiased high-throughput sequencing platform, to detect viral infections in clinical specimens without using any viral genetic information in advance.

In the present study, four patient swab samples were obtained during novel influenza A (H1N1) outbreaks (samples 1, 2 and 3) and seasonal influenza virus A (H3N2) infections (sample 4) in May 2009 from the Laboratory of Beijing CDC, China. One swab sample from a healthy individual (sample 5) was used as a negative control.

Total RNA was isolated from each sample using the QIA-amp RNeasy minikit (Qiagen, Hildenberg, Germany). cDNA was synthesized using the M-MLV RTase cDNA synthesis kit (Takara Bio Inc., Otsu, Japan), and fragmented by nebulizers (Illumina, Inc.) to <800 bp. The overhangs resulting from

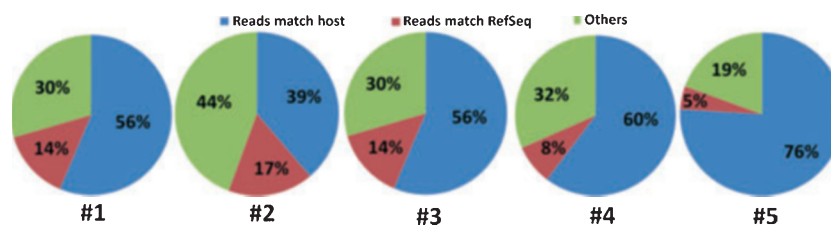
fragmentation were converted into blunt ends using T4 DNA polymerase and *Escherichia coli* polymerase I Klenow fragment. After an 'A' base was added to the 3' end, the adapters were ligated to the ends of the DNA fragments. Unligated adapters were then removed, after which the genomic DNA library was obtained by PCR (primer F: 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCT-3'; primer R: 5'-CAAGCA GAAGACGGCATAACGAGCTCTTCCGATCT-3'). Sequencing and sequence analysis were performed using the SBS sequencing kit v3 and the Genome Analyzer, respectively (Illumina, Inc.).

The quality of the sequencing reads from each sample was ensured by the parallel sequencing of in-house samples, and also included the removal of reads with unrecognized sites or sequencing adaptors, and duplicate reads. Valid reads were screened for host contaminants against the human reference genome using ELAND software, a part of the Solexa analysis pipeline [Cox AJ, unpublished data]. The remaining reads were searched among the microbes subset (viruses, bacteria, fungi and protozoa) of the RefSeq database [4] for taxonomy classification using MAQ software [5].

A summary of the sequencing data obtained using the Solexa system is shown in Fig. 1. The relative ratio of reads matched with the RefSeq database to the total valid reads was 5–17% for each sample. As expected, the ratio of reads matched with human mRNA was in the range 39–56% because total RNA, and not just pathogenic RNA, was isolated from the swab samples. The other 19–44% reads matched with neither host mRNA, nor the RefSeq database, most likely because the RefSeq database mainly focuses on representative strains of taxonomically established organisms.

Sequences corresponding to hundreds of microbes in the swab samples were characterized directly from the sequencing reads matched with the RefSeq database. The fully-characterized microbes are shown in Table 1. The ratio of bacteria was the highest at 63.48–89.65% of total valid reads per sample. The ratio of fungi, protozoa and viruses were between 2.14–14.34%, 3.57–31.71% and 0.13–0.53%, respectively.

Sequence analysis of all identified viral reads was performed to characterize any additional sequences. As shown in Table 1, a BLAST search indicated that 44 (sample 1), 32 (sample 2) and 102 (sample 3) reads of the three samples were novel A (H1N1)-derived, strongly indicating that these patients were similarly infected with novel A (H1N1) virus, consistent with previous RT-PCR diagnostic results (Genebank accession number: GQ183617–GQ183624) [6]. Similarly, 254 reads of sample 4 were derived from seasonal A



**FIG. 1.** Summary of sequencing reads using the Solexa system. Valid reads were used for BLASTN search analyses after removing duplicate reads and reads with unrecognized sites or sequencing adaptors.

**TABLE 1.** Summary of microbe sequence analysis in swab samples

Sample number	1	2	3	4	5
Total reads	298 249	331 249	580 424	176 489	158 327
Fungi	42 779 (14.34%)	21 572 (6.51%)	58 245 (10.04%)	7655 (4.34%)	3396 (2.14%)
Bacteria	239 804 (80.41%)	296 948 (89.65%)	485 943 (83.72%)	112 037 (63.48%)	131 617 (83.13%)
Protozoa	14 733 (4.94%)	11 820 (3.57%)	35 489 (6.11%)	55 970 (31.71%)	22 476 (14.20%)
Virus	933 (0.31%)	909 (0.27%)	747 (0.13%)	827 (0.47%)	838 (0.53%)
Influenza A viruses	44	32	102	254	0
Novel A (H1N1)	44	32	102	0	0
Seasonal A (H3N2)	0	0	0	254	0
Adenovirus	130	162	57	55	80
Herpesvirus	34	32	53	23	18
Murine leukaemia virus	32	28	23	37	36
Mapping to influenza A virus	10.39	8.39	17.79	44.56	0
Genome cover rate (%)					

(H3N2) virus according to BLAST searches. Influenza virus was not detected in the control sample (sample 5). In addition, several common human virus-derived sequences, including adenovirus and herpesvirus were detected in all samples [7]. Maybe as a result of the addition of M-MLV reverse transcriptase during samples preparation, sequences corresponding to murine leukaemia virus were detected.

Nucleic acid sequence-based methods have been used extensively for the diagnosis of viral infection in recent years. However, most methods depend on prior knowledge of pathogen sequences, rendering them ineffective for detecting unexpected or mutated viral sequences. This limitation may be overcome by the use of new high-throughput sequencing strategies. In the present study, we demonstrate the potential of our unbiased pangenomic approach for identifying pathogenic viruses directly without advance genetic information. cDNAs, as templates for the Solexa platform, were prepared by random primer using RNAs extracted from clinical samples. The system produced >5.908 million reads per run within 48 h and enabled us to obtain an almost complete spectrum of nucleotide information from each sample, including bacteria, fungi, protozoa and viruses. Upon further analysis of identified viral reads, influenza sequences were present in 32–254 of the 158 327–580 424 valid reads in each sample (Table 1). Importantly, the coverage rates were in the range 8.39–44.56%, which is sufficient for viral subtype identification. Novel A (H1N1) and seasonal A (H3N2) viruses were detected in the absence of prior genetic infor-

mation, consistent with the classical RT-PCR diagnostic method. This suggests that this approach can contribute to the detection of unexpected or mutated virus by direct comparison with mutant and wild-type viral sequences in the RefSeq database. Likewise, pathogens eliciting similar symptoms could be distinguished in a single sample through analysis of unbiased high-throughput sequencing results, enabling the accurate diagnosis of mixed infection.

## Transparency Declaration

The study was supported in part by grants from the 'AIDS and viral hepatitis and other major infectious diseases prevention and control' technology projects (2009ZX10004-102), the National Basic Research Program of China (2010CB534003 and 2005CB522905), as well as by an intramural grant from the Institute of Pathogen Biology (2008IPB010). The authors declare that there are no competing interests.

## References

- Mahony JB. Detection of respiratory viruses by molecular methods. *Clin Microbiol Rev* 2008; 21: 716–747.
- Hervás F. Chip-mediated techniques: how close are we to generalised use in the infectious disease clinic? *Clin Microbiol Infect* 2004; 10: 865–867.

3. 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: e4219.
4. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 2007; 35(Database issue): D61–D65.
5. Heng L, Jue R, Richard D. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 2008; 18: 1851–1858.
6. Yang J, Yang F, Huang F, Wang J, Jin Q. Subclinical infection with the novel Influenza A (H1N1) virus. *Clin Infect Dis* 2009; 49: 1622–1623.
7. Ren L, Gonzalez R, Wang Z *et al*. Prevalence of human respiratory viruses in adults with acute respiratory tract infections in Beijing, 2005–2007. *Clin Microbiol Infect* 2009; 15: 1146–1153.

## First detection of group C rotavirus in children with acute gastroenteritis in South Korea

S. Moon<sup>1,3</sup>, C. D. Humphrey<sup>2</sup>, J. S. Kim<sup>3</sup>, L. J. Baek<sup>3</sup>, J.-W. Song<sup>3</sup>, K.-J. Song<sup>3</sup> and B. Jang<sup>1</sup>

1) *Gastroenteritis and Respiratory Viruses Laboratory Branch and 2) Infectious Disease Pathology Branch, Centers for Disease Control and Prevention, Atlanta, Georgia, USA and 3) Department of Microbiology and Institute for Viral Diseases, College of Medicine, Korea University, Seoul, South Korea*

### Abstract

Group C rotavirus (GpC RV) causes sporadic cases and outbreaks of acute diarrhoea in humans worldwide, but has not been detected among children in South Korea. The present study aimed to detect GpC RV among children hospitalized with gastroenteritis in South Korea and to perform a molecular characterization of GpC RV strains. From November 2003 to January 2006, 434 faecal samples were collected from children <10 years of age who were hospitalized for treatment of acute diarrhoea and screened for group C and A rotaviruses by enzyme immunoassay. GpC RV strains were characterized by sequence and phylogenetic analysis. Of the 434 samples screened, two were positive for GpC RV and one had a mixed GpC and GpA RV infection. One of the strains, Icheon, shared high sequence conservation in VP4, VP6 and VP7 genes with other published GpC RV. This is the first report describing the molecular characteristics of GpC RV among children in South Korea. Additional surveillance is needed to determine the burden of GpC RV gastroenteritis.

**Keywords:** Diarrhoea, group C rotavirus, molecular epidemiology, phylogenetic analysis

**Original Submission:** 5 March 2010; **Revised Submission:** 14 April 2010; **Accepted:** 7 May 2010

Editor: J.-M. Pawlowsky

**Article published online:** 19 May 2010

*Clin Microbiol Infect* 2011; 17: 244–247

10.1111/j.1469-0691.2010.03270.x

**Corresponding author:** B. Jang, Division of viral Diseases, MS G04, National Center for Immunization and Respiratory Disease, 1600 Clifton Road, Atlanta, GA 30333, USA  
E-mail: bxj4@cdc.gov

Rotaviruses are the most important cause of acute gastroenteritis in humans and a variety of animals [1]. Rotaviruses possess a segmented, double-stranded RNA genome and are classified into seven groups (A–G) based on their distinctness with respect to genes and capsid proteins. Group A, B and C rotaviruses cause disease in both humans and animals [1]. Group C rotavirus (GpC RV), first detected in piglets [2] and an infant [3], has been associated with sporadic cases and outbreaks of gastroenteritis among children and adults in many countries [4–9]. GpC RV is generally considered to cause a relatively milder disease, with fewer episodes of vomiting per day and less dehydration compared to group A rotavirus (GpA RV) infection in children [4,7].

All human GpC RV strains detected to date throughout the world have shown high sequence conservation and belong to single G and P genotypes [10,11]. By contrast, human GpC strains are divergent from those in cattle and swine [11–14]. In the present study, for the first time, we detected GpC RV in faecal samples from children with acute diarrhoea in South Korea and determined the genetic relatedness of the Korean isolate to other published strains.

A total of 434 stool samples was collected from patients <10 years of age, who were hospitalized for treatment of gastroenteritis at Sungmo Hospital, Icheon, South Korea, from November 2003 to January 2006. All 434 samples were screened for GpA RV by using a Rotatek™ kit (South Korea Green Cross Corp., Seoul, South Korea) [15] and, for GpC RV, by using an in-house immunoassay developed with hyper-immune sera against porcine GpC RV Cowden strain and human GpC RV VLPs [4,16]. Briefly, 96-well plates were coated with porcine anti-GpC serum (U340; dilution 1:2000) or normal porcine serum (Z1329; dilution 1:2000) in coating buffer (35 mM NaHCO<sub>3</sub>, 15 mM Na<sub>2</sub>CO<sub>3</sub>, pH 9.6) and incubated overnight at 4°C. Plates were washed with phosphate buffered saline-Tris (pH 7.6) and incubated with blotto for 1 h at 37°C. Plates were washed and then incubated with diluted stool samples (dilution 1:10) or GpC VLPs (positive