Brief Communication
Clinical Microbiology

Ann Lab Med 2016;36:255-258
http://dx.doi.org/10.3343/alm.2016.36.3.255
ISSN 2234-3806 • eISSN 2234-3814

ANNALS OF LABORATORY MEDICINE

Genotyping Influenza Virus by Next-Generation Deep Sequencing in Clinical Specimens

Moon-Woo Seong, M.D., Sung Im Cho, M.S., Hyunwoong Park, M.D., Soo Hyun Seo, M.D., Seung Jun Lee, M.D., Eui-Chong Kim, M.D., and Sung Sup Park, M.D.

Department of Laboratory Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea

Rapid and accurate identification of an influenza outbreak is essential for patient care and treatment. We describe a next-generation sequencing (NGS)-based, unbiased deep sequencing method in clinical specimens to investigate an influenza outbreak. Nasopharyngeal swabs from patients were collected for molecular epidemiological analysis. Total RNA was sequenced by using the NGS technology as paired-end 250 bp reads. Total of 7 to 12 million reads were obtained. After mapping to the human reference genome, we analyzed the 3-4% of reads that originated from a non-human source. A BLAST search of the contigs reconstructed *de novo* revealed high sequence similarity with that of the pandemic H1N1 virus. In the phylogenetic analysis, the *HA* gene of our samples clustered closely with that of A/Senegal/VR785/2010(H1N1), A/Wisconsin/11/2013(H1N1), and A/Korea/01/2009(H1N1), and the *NA* gene of our samples clustered closely with A/Wisconsin/11/2013(H1N1). This study suggests that NGS-based unbiased sequencing can be effectively applied to investigate molecular characteristics of nosocomial influenza outbreak by using clinical specimens such as nasopharyngeal swabs.

Key Words: Pandemic H1N1 virus, Influenza virus, Nosocomial outbreak, Next-generation sequencing, Genome sequencing

Received: July 8, 2015
Revision received: November 9, 2015
Accepted: December 21, 2015

Corresponding author: Sung Sup Park
Department of Laboratory Medicine, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea
Tel: +82-2-2072-3206
Fax: +82-2-747-0359
E-mail: sparkle@snu.ac.kr

Nosocomial transmission of influenza is a recognized problem for vulnerable patients, including pediatric, immunosuppressed, hematologic-oncologic, and elderly patients [1-3]. Therefore, rapid and accurate identification of an influenza outbreak is essential for patient care and treatment.

Respiratory viral infections are mostly diagnosed on the basis of culture, rapid antigen test, or molecular diagnostic assays. Although molecular diagnostic assays show superior sensitivity than the conventional assays, these are generally designed to identify only certain target viruses [4, 5]. Recently developed next-generation sequencing (NGS) technology can provide deep sequencing in an untargeted manner. Here we demonstrate that NGS-based unbiased deep sequencing can be a useful tool to accurately identify virus subtypes from the clinical specimens in the event of a nosocomial influenza outbreak.

An outbreak of influenza-like illness was suspected in a pulmonary ward of Seoul National University Hospital, Seoul, Korea on Jan 2014. More than 10 individuals, including the ward nurses and patients were evaluated as Flu A positive on the basis of influenza antigen test using a BD Directigen EZ Flu A+B Test kit (BD Diagnostics, Sparks, MD, USA) on 20-22 Jan, 2014. Subsequently, two molecular studies were simultaneously performed for six patients, whose specimens were available: 1) a conventional study using respiratory virus multiplex PCR and seasonal H1/H3 PCR, and 2) an NGS-based whole genome sequencing. For the respiratory virus multiplex PCR, we used the Seeplex RV 12 ACE Detection kit (Seegene, Seoul, Korea) according to the manufacturer's instructions. The seasonal H1/H3 PCR was carried out by using Seeplex FluA ACE Subtyping kit (Seegene). This study was approved by the Institutional Review Board of Seoul National University Hospital.

For whole viral genome sequencing, total RNA was extracted

by using QIAamp Viral RNA Mini kit (Qiagen, Hilden, Germany). Double-stranded cDNA was prepared from 5 μg of total RNA by the random priming method using the SuperScript Choice System for cDNA synthesis (Invitrogen, Carlsbad, CA, USA), and processed with TruSeq Stranded Total RNA Sample Prep kit (Illumina Inc., San Diego, CA, USA) according to the manufacturer's instruction. The resulting library was quantified and assessed for its quality by using Bioanalyzer 2100 (Agilent, Palo Alto, CA, USA), and the average library size was 300 bp. This library was sequenced by using Illumina MiSeq (Illumina Inc.) with reagent kit v2 (Illumina Inc.). Sequencing was performed with paired-end reads of 250 bp in length.

Bioinformatic analyses were carried out after the procedures. First, total reads were mapped to the human hg19 reference genome to filter out reads that originated from the human genome, and unmapped reads were collected for further analysis by using a mapping module in CLC Genomics Workbench software version 6.5.1 (CLC bio, Cambridge, MA, USA). Second, the genome assembly was constructed from these unmapped reads by using a *de novo* module in the same software, with minimum contiguous length set at 200 bp for assembling consensus sequences and the other parameters set at default. Third, the assembled contigs (≥500 bp) were compared against the nucleotide database of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) by using the BLAST algorithm (http://blast.ncbi.nlm.nih.gov/). The highest scoring BLAST match was filtered according to a minimum 90% identity and 90% query coverage. Then, common BLAST matches were selected for phylogenetic analysis. Finally, assembled contigs and candidate viral reference genome sequences were aligned in ClustalW (http://www.clustal.org), and phylogenetic analysis was performed by using the maximum likelihood method.

The Seeplex RV 12 PCR confirmed influenza A in three of these patients (P1, P4, and P6), but failed to demonstrate influenza A in the remaining three patients (P2, P3, and P5). Seasonal H1 and H3 were excluded for the influenza A positive patients. The molecular study results are summarized in Table 1.

Total of 7 to 12 million reads were obtained from five patients. The sample from patient P5 was excluded from whole-genome sequencing because the extracted RNA quality was poor (Table 2). After mapping to the human reference genome, we further analyzed the 3-4% of total reads that were not identified as originating from a human source.

With *de novo* assembly of nonhuman reads, an average of 14,999 contigs per patient was constructed. We selected 5,956 total contigs with a minimum length of 500 bp and performed a BLAST search to identify similar viral genomes (Table 2).

Our BLAST search indicated that most contigs were originated from human and bacterial (e.g., *Staphylococcus epidermidis* and *Pseudomonas* spp.) sources. The remaining contigs showed high similarity to pandemic H1N1 virus sequences. The size of these contigs ranged from 524 bp to 2,299 bp (average 844 bp).

Table 1. Molecular study results for six patients*

Patient	Multiplex RV PCR	Seasonal H1/H3 PCR	*HA* typing by WGS†	*NA* typing by WGS†
P1	Flu A	Negative	PDM H1N1/09	PDM H1N1/09
P2	Negative	Negative	PDM H1N1/09	-‡
P3	Negative	Negative	-‡	-‡
P4	Flu A	Negative	PDM H1N1/09	PDM H1N1/09
P5	Negative	Negative	-§	-§
P6	Flu A	Negative	PDM H1N1/09	PDM H1N1/09

*All patients were Flu A positive by BD Directigen EZ Flu A+B Test kit; †Whole-genome sequencing (WGS) results are based on contigs ≥500 bp only; ‡BLAST search for the *HA* and *NA* genes failed in these cases; §P5 was excluded from WGS because the quality of extracted RNA was poor.

Table 2. Sequencing statistics of the study

Patient	Total reads (%)	Human reads (%)	Contigs ≥500 bp	H1N1 reads (%)*	H1N1 coverage	*HA* reads (%)†	*NA* reads (%)†
P1	8,539,246 (100)	8,344,459 (97.7)	1,059	382 (0.0045)	6.2×	64 (16.8)	60 (15.7)
P2	7,717,044 (100)	7,534,324 (97.6)	1,294	216 (0.0028)	4.2×	27 (12.5)	31 (14.4)
P3	8,406,782 (100)	8,228,729 (97.9)	1,473	25 (0.0003)	1.2×	7 (28.0)	6 (24.0)
P4	12,038,410 (100)	11,649,779 (96.8)	1,859	1,187 (0.0099)	14.8×	109 (9.2)	168 (14.2)
P5‡	-	-	-	-	-	-	-
P6	1,526,958 (100)	1,474,422 (96.6)	271	1,207 (0.079)	14.7×	93 (7.7)	123 (11.0)

*The pandemic H1N1 reference genome used here is the A/Korea/01/2009(H1N1) strain. GenBank accessions are GQ160811-3, GQ131023-6, and GQ132185; †These percentages are for reads only mapped to the pandemic H1N1 reference genome; ‡P5 was excluded from whole-genome sequencing because the quality of extracted RNA was poor.

http://dx.doi.org/10.3343/alm.2016.36.3.255

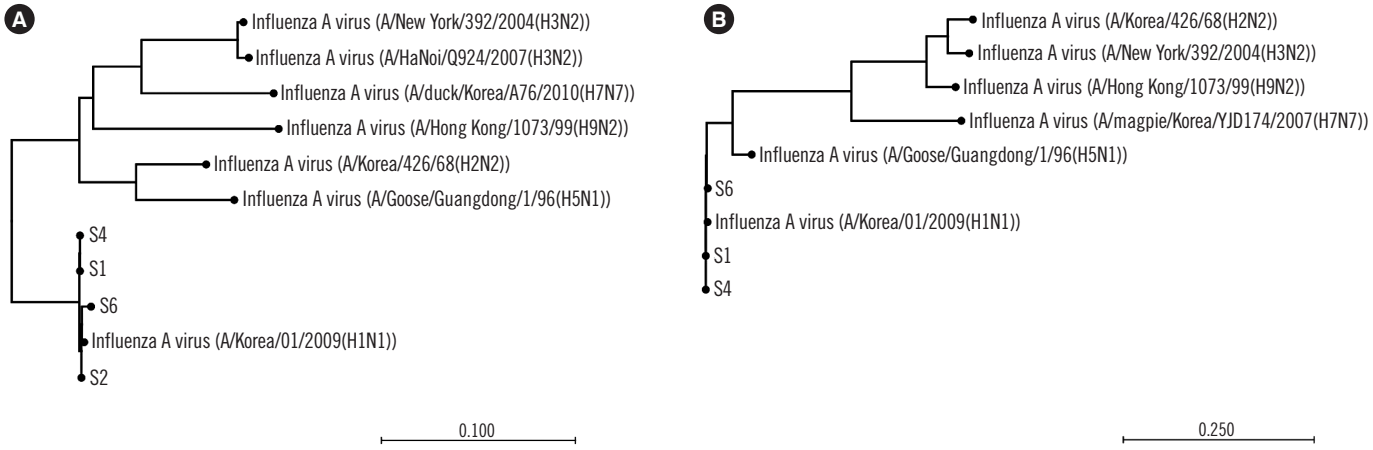

Fig. 1. Phylogenetic analysis of the *HA* (A) and *NA* (B) genes.

Phylogenetic analysis of the *HA* and *NA* genes included various influenza A strains like A/Korea/01/2009(H1N1), A/Korea/426/68(H2N2), A/New York/392/2004(H3N2). The analysis showed that sequences from both the *HA* and *NA* genes of our samples clustered closely with sequences from the virus strain A/Korea/01/2009(H1N1) (Fig. 1): percent identity between our samples and A/Korea/01/2009(H1N1) was 98.6-99.6% for *HA* and 99.2-99.5% for *NA*.

We analyzed the genome coverage of the non-human reads for the influenza A(H1N1)pdm09 virus using A/Korea/01/2009 (H1N1) (GenBank accessions: GQ160811-3, GQ131023-6, GQ 132185), one closely related pandemic H1N1 strain for which the whole genome sequence is available. An average of 603 reads (25-1,207) mapped to this pandemic H1N1 virus and the overall read coverage was 8.2× (1.2×-14.8×) (Table 2). Among the reads mapped to this pandemic virus, 7.7-28.0% reads were mapped to the *HA* gene, and 11.0-24.0% were mapped to the *NA* gene.

Whole-genome sequencing of influenza A virus has been used to determine the genetic basis of pathogenicity and antiviral resistance and to identify mixed infections or quasispecies [6, 7]. Whole-genome sequencing has also been applied for various epidemiological investigations, such as outbreaks of neonatal methicillin-resistant *Staphylococcus aureus*, *Mycobacterium tuberculosis*, and multi-drug resistant *Escherichia coli* [8-10]. To our knowledge, however, this study is the first to investigate an influenza outbreak by whole-genome sequencing in clinical specimens. Unlike other molecular methods that can detect only a limited number of virus targets, whole-genome sequencing can provide information unbiased by prior knowledge of the viral etiology of an outbreak. In addition, deep sequencing enables us to identify the exact cause of an epidemic event in clinical specimens with viral RNA or DNA in very small quantities, for example, in specimens containing viral targets equivalent to 0.005% of the total sequencing reads as seen in our study.

As well as methodological robustness and cost, the turnaround time (TAT) needs to be considered before whole-genome sequencing is applied to the investigation of nosocomial outbreaks by a diagnostic microbiology laboratory. Sherry *et al*. [8] reported five days of TAT from a positive culture to the completion of sequencing to investigate putative multidrug-resistant *E. coli*. In our study, the TAT from library preparation for MiSeq sequencing to completion of sequence analysis was three days: two days for sequencing and one day for sequence analysis. This time is longer than that of other molecular methods, but this time might be improved with the advancement of sequencing technologies. The value of unbiased information at the sequence level, which this approach provides, should also be considered when choosing a method of investigation.

Until now, at least seven phylogenetically distinct viral clades of pandemic H1N1 virus have been identified [11]. The pandemic H1N1 virus identified in our study was most closely related to the A/Wisconsin/11/2013 strain. Among the amino acid changes defining clades 1-7, this virus had unique amino acid substitutions in the NP (V100I), NA (N248D), NS1 (I123V), and HA (S220T) regions. Although there were no substitutions at codon 106 of NA, this virus can be classified as clade 7, the most commonly isolated clade of H1N1 influenza virus in the world [11].

In summary, NGS-based unbiased sequencing can be effectively applied to investigate molecular characteristics of nosocomial influenza outbreak in clinical specimens such as nasopharyngeal swabs.

Authors' Disclosures of Potential Conflicts of Interest

No potential conflicts of interest relevant to this article were reported.

Acknowledgments

This research was supported by funding (code 2013E4700100) from the Research of Korea Centers of Disease Control and Prevention.

REFERENCES

1. Bearden A, Friedrich TC, Goldberg TL, Byrne B, Spiegel C, Schult P, et al. An outbreak of the 2009 influenza a (H1N1) virus in a children's hospital. Influenza Other Respir Viruses 2012;6:374-9.
2. Pollara CP, Piccinelli G, Rossi G, Cattaneo C, Perandin F, Corbellini S, et al. Nosocomial outbreak of the pandemic Influenza A (H1N1) 2009 in critical hematologic patients during seasonal influenza 2010-2011: detection of oseltamivir resistant variant viruses. BMC Infect Dis 2013; 13:127.
3. Tsagris V, Nika A, Kyriakou D, Kapetanakis I, Harahousou E, Stripeli F, et al. Influenza A/H1N1/2009 outbreak in a neonatal intensive care unit. J Hosp Infect 2012;81:36-40.
4. Mengelle C, Mansuy JM, Da Silva I, Guerin JL, Izopet J. Evaluation of a polymerase chain reaction-electrospray ionization time-of-flight mass spectrometry for the detection and subtyping of influenza viruses in respiratory specimens. J Clin Virol 2013;57:222-6.
5. Munro SB, Kuypers J, Jerome KR. Comparison of a multiplex real-time PCR assay with a multiplex Luminex assay for influenza virus detection. J Clin Microbiol 2013;51:1124-9.
6. Ramakrishnan MA, Tu ZJ, Singh S, Chockalingam AK, Gramer MR, Wang P, et al. The feasibility of using high resolution genome sequencing of influenza A viruses to detect mixed infections and quasispecies. PLoS One 2009;4:e7105.
7. Renzette N, Caffrey DR, Zeldovich KB, Liu P, Gallagher GR, Aiello D, et al. Evolution of the influenza A virus genome during development of oseltamivir resistance in vitro. J Virol 2014;88:272-81.
8. Sherry NL, Porter JL, Seemann T, Watkins A, Stinear TP, Howden BP. Outbreak investigation using high-throughput genome sequencing within a diagnostic microbiology laboratory. J Clin Microbiol 2013;51:1396-401.
9. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis 2013; 13:137-46.
10. Köser CU, Holden MT, Ellington MJ, Cartwright EJ, Brown NM, Ogilvy-Stuart AL, et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N Engl J Med 2012;366:2267-75.
11. Nelson M, Spiro D, Wentworth D, Beck E, Fan J, Ghedin E, et al. The early diversification of influenza A/H1N1pdm. PLoS Curr 2009;1: RRN1126.

http://dx.doi.org/10.3343/alm.2016.36.3.255