Journal of Clinical Virology 73 (2015) 84–88

Contents lists available at ScienceDirect

Journal of Clinical Virology

journal homepage: www.elsevier.com/locate/jcv

Rapid genome sequencing and characterization of novel avian-origin influenza A H7N9 virus directly from clinical sample by semiconductor sequencing

Yongfeng Hu[a,1], Xianwen Ren[a,1], Li Li[a], Yan Xiao[a], Jie Dong[a], Lilian Sun[a], Yafang Zhu[a], Fan Yang[a,*], Xi Zhang[c,*], Qi Jin[a,b,*]

[a] MOH Key Laboratory of Systems Biology of Pathogens, Institute of Pathogen Biology, CAMS & PUMC, Beijing 100176, PR China
[b] Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Hangzhou, PR China
[c] Shanghai Municipal Center for Disease Control and Prevention, Shanghai, PR China

ARTICLE INFO

Article history:
Received 7 July 2015
Received in revised form 14 October 2015
Accepted 26 October 2015

Keywords:
Genome sequencing
Influenza A H7N9 virus
Clinical sample
Semiconductor sequencing

ABSTRACT

Background: Recent outbreaks of severe pneumonia or acute respiratory distress syndrome have attracted much public interest. Rapid and accurate diagnosis of the causative agent is key for an adequate response to suspected outbreaks.
Objectives: We report a case that highlights the potential of semiconductor sequencing to rapidly determine the novel virus genome sequences.
Study design: We have developed a method for rapid *de novo* assembly of the novel influenza A H7N9 virus genome directly from the tracheal aspirate of a patient using semiconductor sequencer without culture and prior sequence information. Further, characteristic amino acids were analyzed and phylogenetic analysis were done for key genes of the influenza A virus.
Results: Deep sequencing yielded 435,239 reads assigned to H7N9 viruses, with an average length of 172 bp, accounting for 18.6% of total reads (2,339,680). Complete genome of the virus was obtained by *de novo* assembly method within 2 days. Genomic average depth of coverage of the Ion Torrent PGM was up to 5679 fold. Selected characteristic amino acids were observed, and phylogenetic analyses showed that the novel H7 virus was genetically close to 2011 duck H7N3 viruses in Zhejiang. The novel N9 sequences were most closely related to gene sequences of N9 derived from ducks H11N9 in 2011 in Jiangxi and H2N9 sequences from Hong Kong in 2010, in China, and therefore they may share a common ancestor.
Conclusions: The sequence-independent semiconductor sequencing is a powerful tool to investigate outbreak of a novel pathogen.

1. Background

Emerging of a novel avian-origin influenza virus H7N9 marked by severe pneumonia and high mortality in China, is a significant public health issue and is of great concern [1]. Humans are rarely infected with avian influenza viruses. Severe disease in humans caused by a novel influenza A virus might be sporadic human infections from an animal source, or signal the start of an new influenza pandemic caused by the established virus [2].

Accurate diagnosis is critical for pandemic influenza recognition, surveillance, and public health interventions. Availability of influenza virus sequences and phylogenetic studies have largely contributed to a better understanding of the emergence, spread and evolution of influenza virus epidemics. The next-generation sequencing (NGS), without culture and prior knowledge of sequences, is especially attractive in the study of novel virus that are too divergent from known viruses to be detected by either PCR or microarray techniques [1]. The Ion Torrent PGM is a semiconductor sequencing technology and has a advantage over other next-generation sequencers in speed, operation, and cost which is widely used as a diagnostic in the health care industry [3,4].

* Corresponding authors at: No. 6, Rongjing east street, BDA, Beijing, PR China. Fax: +86 10 6787 7736.
E-mail addresses: yangfan@ipbcams.ac.cn (F. Yang), xzhang@scdc.sh.cn (X. Zhang), zdsys@sina.com (Q. Jin).
[1] These authors contributed equally to this study.

http://dx.doi.org/10.1016/j.jcv.2015.10.022
1386-6532/

2. Objectives

We reported a case for rapid determination of full genome of the novel avian-origin Influenza A (H7N9) virus directly from a tracheal aspirate sample of one patient using semi-conductor sequencing technology.

3. Study design

3.1. Patient and sample collection

A 74-year-old man began to get fever in March, 2013. He sought medical attention when fevers continued, and was admitted to hospital because of severe pneumonia and ARDS (Acute Respiratory Distress Syndrome). Endotracheal intubation and mechanical ventilator support was given because of respiratory failure. A tracheal aspirate specimen was collected into transport medium on day 8 after symptom onset through endotracheal tube suctioning [5,6], and sent to the diagnostic laboratory within two hours. Original clinical sample was subtyped as H7N9 with real-time RT-PCR provided by WHO Collaborating Centre in Beijing.

This study was reviewed and approved by the Ethics Committee of the Institute of Pathogen Biology, CAMS & PUMC. Written informed consent was obtained for the use of tracheal aspirate samples from the patient's relatives.

3.2. Viral enrichment and RNA purification

A total of 200 μl tracheal aspirate was first filtered through a 0.45-μm-pore-size polyvinylidenedifluoride (PVDF) filter (Millipore) to remove eukaryotic cell- and bacterium-sized particles, and then digested in a cocktail of DNase and RNase enzymes consisting of 14 U of Turbo DNase (Ambion), 20 U of benzonase (Novagen) and 20 U of RNase One (Promega) at 37 °C for 2 h in DNase buffer (Ambion) [7]. The RNA of treated tracheal aspirate was purified with a QIAamp viral RNA minikit (Qiagen) (Fig. 1).

3.3. Whole transcription amplification

First-strand cDNA was synthesized with 100 pmol of primer K-8N (GACCATCTAGCGACCTCCAC-NNNNNNNN), as previously described [8]. Double-stranded cDNA were produced with Klenow fragment (NEB). Sequence-independent PCR amplification was conducted with 5 μl of the double-stranded cDNA, 1 μl Phusion HF buffer, 200 μM deoxynucleoside triphosphate (dNTP), 1 μM primer K (GACCATCTAGCGACCTCCAC), and 0.5 U Phusion DNA polymerase (NEB). The PCR cycling was performed as follows: 98 °C for 30 s, followed by 35 cycles of 98 °C for 10 s, 55 °C for 30 s, and 72 °C for 1 min, with a final extension at 72 °C for 10 min [8]. A DNA smear larger than 500 bp was excised and purified with a Min-Elute gel extraction kit (Qiagen), and used for deep sequencing.

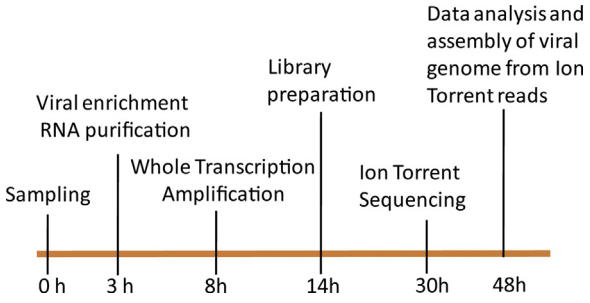

Fig. 1. Timeline of viral detection and full genome sequencing by Ion Torrent PGM.

3.4. Deep sequencing and data analysis

We used the NEBNext® Fast DNA Library Prep Set (NEB) to prepare library and a range of 290–330 bp was selected. Sequencing was performed using Ion Torrent PGM (Life Technologies, Guilford, CT, USA) with 316 chip and 200 bp sequencing chemistry.

All reads were aligned to the NCBI non-redundant nucleotide database (nt) by the BLAST software (version 2.2.22) after filtering adapters and human-origin reads. Alignments of the reads with scores >80 were retained for the subsequent analysis. The reads assigned to H7N9 were subjected to *de novo* contig assembly using the Lasergene program SeqMan Pro, with 80% minimum sequence similarity tolerance and 30 bp minimum match size. Reference-based mapping method was applied to the data for assembly into viral genome. The gaps between contigs were filled up by PCR based on specific primers designed from contig sequences, and the 5′and 3′ ends of eight genome segments were completed with a 5′RACE and a 3′RACE kit (Invitrogen). Full genome sequences of the virus (A/Shanghai/ION/2013) was deposited in GenBank under accession numbers KP969031–KP969038.

3.5. Phylogenetic analysis

We downloaded H7N9 virus genome sequences from the Global Initiative on Sharing Avian Influenza Data (GISAID) database, and public sequences from the Influenza Virus Resource at NCBI [9]. Providers of sequences downloaded from GISAID, listed with accession numbers, are acknowledged in Table S1.

Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 6 [10]. We performed multiple sequence alignments with the Muscle program of MEGA software. Phylogenetic trees were inferred from the sequence alignments on the basis of maximum likelihood methods, the general time-reversible model of nucleotide substitution, and γ-distributed rates among sites to estimate the viral gene relationship with selected influenza A virus strains. The robustness of the tree topology was assessed by bootstrap analysis of 1000 pseudo-replicates of the sequences.

4. Results

4.1. Determination of the pathogen—the novel influenza A H7N9 virus

A summary of the Ion Torrent sequencing data obtained is shown in Table 1. Deep sequencing of the sample generated 2,339,680 valid reads, with an average read length of 172 bp. A BLAST search showed that a total of 435,239 reads (accounting for 18.6% of total reads and 97.6% of viral reads) from the patient were novel influenza A (H7N9)-derived, strongly indicating that the patient was infected with the novel influenza A (H7N9) virus. Meanwhile, the specimen was detected by traditional real-time PCR (RT-PCR) specific for influenza virus A (H1, H3), H5N1 and H7N9, or influenza virus B. The patient was later identified

Table 1
Data summary of Ion Torrent sequencing of a tracheal aspirate sample from the patient with influenza A(H7N9) virus infection.

Total reads	2,339,680
Read length (bp)	172
Cellular organism reads	1,378,652
Viral reads	445,857
Influenza virus (H_XN_Y)	435,239 (H7N9)
Average depth of coverage	5679
RT-PCR (IF-A/H7N9)	+/+