Original Article

Evaluation of A Single-reaction Method for Whole Genome Sequencing of Influenza A Virus using Next Generation Sequencing*

ZOU Xiao Hui, CHEN Wen Bing, ZHAO Xiang, ZHU Wen Fei, YANG Lei, WANG Da Yan, and SHU Yue Long**

Chinese National Influenza Center, National Institute for Viral Disease Control and Prevention; China CDC, Key Laboratory for Medical Virology, Ministry of Health, Beijing 102206, China

Abstract

Objective To evaluate a single-reaction genome amplification method, the multisegment reverse transcription-PCR (M-RTPCR), for its sensitivity to full genome sequencing of influenza A virus, and the ability to differentiate mix-subtype virus, using the next generation sequencing (NGS) platform.

Methods Virus genome copy was quantified and serially diluted to different titers, followed by amplification with the M-RTPCR method and sequencing on the NGS platform. Furthermore, we manually mixed two subtype viruses to different titer rate and amplified the mixed virus with the M-RTPCR protocol, followed by whole genome sequencing on the NGS platform. We also used clinical samples to test the method performance.

Results The M-RTPCR method obtained complete genome of testing virus at 125 copies/reaction and determined the virus subtype at titer of 25 copies/reaction. Moreover, the two subtypes in the mixed virus could be discriminated, even though these two virus copies differed by 200-fold using this amplification protocol. The sensitivity of this protocol we detected using virus RNA was also confirmed with clinical samples containing low-titer virus.

Conclusion The M-RTPCR is a robust and sensitive amplification method for whole genome sequencing of influenza A virus using NGS platform.

Key words: Influenza A virus; Whole genome sequencing; NGS

INTRODUCTION

Influenza virus is one of the most common agents causing pervasive infections worldwide. While seasonal influenza virus usually causes mild respiratory illnesses, novel influenza viruses of zoonotic origin (mainly avian influenza virus) sporadically cross host barrier and cause severe human infections[1]. The high mutation rate of the viral RNA polymerase makes influenza virus genotype highly variable in the natural ecosystem. Influenza virus has also evolved

---

*This work was funded by a project (2014ZX10004002) of the Chinese National Key Program of Mega Infectious Disease of the National 12th Five-Year Plan.

**Correspondence should be addressed to SHU Yue Long, Professor, PhD, Tel/Fax: 86-10-58900850, E-mail: yshu@cnic.org.cn

Biographical note on first author: ZOU Xiao Hui, male, born in 1987, Doctoral degree candidate, majoring in pathogen biology.
through genetic reassortment, in which gene segments are exchanged among distinct viruses. Consequently, a highly variable composition of influenza genome has been produced in nature and further leads to different phenotypes. Therefore, characterization of the whole genome is critical for risk evaluation of certain influenza virus to public health.

The first step of influenza virus whole genome sequencing (WGS) is to reverse-transcribe and amplify the RNA genome. So far several protocols have been developed for this purpose. Most of these methods took advantage of the conserved 12 nucleotides at the 3' terminus and 13 nucleotides at the 5' terminus of the vRNA. Hoffmann[2] first introduced a universal primer set for the full-length amplification of all influenza A viruses, which used several segment-specific primers to generate full-length cDNAs. Zhou et al.[3] and Bourret et al.[4] further applied a single reaction to amplify the eight segments of influenza A virus. Hoper[5] developed a different strategy for full-length genome sequencing of influenza A virus using the next generation sequencing (NGS) platform 454FLX. With this method, every genome segment was divided into two amplicons covering its full length and subjected to further sequencing.

The advance in NGS technology has greatly improved our ability to launch a large scale of WGS of influenza virus. With more full-genome data, we can conduct more intensive and comprehensive phylogenetic analysis and achieve a deeper understanding in the epidemiology of influenza virus. In fact, NGS has already been applied in WGS of large amounts of influenza A virus[6]. Currently, a simple and robust amplification of the entire genome of influenza A virus is the prerequisite for successful WGS using NGS. Thus, the method developed by Zhou[3], a multisegment reverse transcription-PCR (M-RTPCR) approach that simultaneously amplified eight genomic RNA segments of influenza A virus, is very suitable to apply for its simplicity (single reaction) and universality (full-subtype amplification).

In this study, we evaluated the sensitivity of M-RTPCR for WGS of influenza A virus and its ability to differentiate mix-subtype virus with biased titer. We also verified the performance of M-RTPCR by directly sequencing the influenza A genome in clinical samples. Our study provided metrics for other researchers to conduct WGS of influenza A virus using NGS.

**MATERIALS AND METHODS**

**Virus and Virus Quantification**

The H1N1 influenza virus A/Puerto Rico/8/34 (PR8) and A/Quail/Hong Kong/G1/97 (H9N2, G1) used in this study were reverse-genetically rescued with PHW-2000 plasmid system to assure their purity[7]. Virus genome was extracted with a RNeasy mini kit (QIAGEN, Hilden, Germany) and separated into aliquots saved in -80 °C. The quantification of virus copy was performed using a standard curve generated by cycle threshold values obtained from serial 10-fold dilutions of in vitro transcripts containing $10^3$-$10^8$ copies of the full-length M genes from influenza A virus.

**Virus RNA mixing and M-RTPCR**

In order to evaluate the method sensitivity of WGS for single subtype virus, we serially diluted PR8 RNA to 5, 25, and 125 copies/μL, from which 5 μL was subjected to M-RTPCR[8]. For evaluation of its ability to differentiate mixed virus with biased copies, four libraries were established. Each library contained 125 copies/μL of PR8 RNA, to which an expected to be detected in the background of the predominant virus (G1) with titers 1000-, 500-, 200-, and 100-fold higher than PR8, respectively. Therefore, the low-titer virus (PR8) was expected to be detected in the background of the equal volume of 125,000, 62,500, 25,000, and 12,500 copies/μL of G1 RNA were spiked in, respectively. The mixed RNAs were reverse-transcribed and amplified with the same M-RTPCR method. The obtained cDNA was quantified with a Qubit® dsDNA HS Assay Kit (Life Technologies, New York, USA). All cDNA products were stored at -20 °C until use.

**Next Generation Sequencing**

All of the cDNA obtained from PR8 and G1 was subjected to NGS on the Ion Torrent PGM platform. Briefly, 100 ng cDNA was sheared with an Ion Shear kit and ligated with P1 and Xpress barcode adaptor using the IonXpress™ Plus gDNA Fragment Library kit. Fragment distribution was checked with the Agilent 2100 Bioanalyzer using a High Sensitivity Chip (Agilent Technologies, Santa Clara, USA). For sensitivity evaluation of single subtype virus, libraries from three PR8 RNA dilutions were pooled at equal mass and subject to emulsion-PCR using the Ion PGM™ Template OT2 200 Kit. For sensitivity of differentiating mixed virus, libraries from four RNA mixtures were pooled and emulsion-PCR as forward.
Library dilution factor and pooling style was determined from a qPCR standard curve using an Ion Library Quantification Kit. Recovered template-positive Ion sphere particles (ISPs) were subjected to quality control with an Ion Sphere Quality Control Kit using a Qubit® 2.0 fluorometer. After enrichment with Ion PGM™ Enrichment Beads, samples containing an optimum number of template ISPs were subjected to standard 200 bp sequencing on the 316v2 chip using a Ion PGM™ Sequencing 200 Kit v2. The reagents for NGS were purchased from Life Technologies.

**Clinical Sample Sequencing**

We collected four pharyngeal swab specimens from patients confirmed to be influenza A positive by RT-PCR targeting the influenza A virus M segment. The specimens were then processed for RNA extraction with a QIAGEN RNeasy mini kit. The RNA samples were reverse-transcribed, amplified, and quantified with the method described above. Samples with sufficient cDNA produced were subjected to PGM sequencing, as described above. For samples with low-titer virus, and thus too little cDNA product obtained for PGM library preparation, we used the Illumina Nextera XT DNA library preparation method (Illumina, San Diego, CA, USA) to prepare sequencing libraries and sequence them on a MiSeq platform with a PE500 read length.

**NGS Data Processing**

NGS data analysis was conducted with CLC genomics workbench V7.5 software. Briefly, reads with a mean length of approximately 200 bp were quality-trimmed to remove reads with an error rate >0.05 or a length <100 bp. The reads obtained from each dilution and mixture were normalized to an equal amount in each group. Reads mapping was performed in CLC Reference Mapper with default settings, but adjusted Length fraction=0.9 and Similarity fraction=0.9. G1 and PR8 whole genomes deposited in NCBI were used as references for mapping (Supplementary data, Reference for mapping). The coverage depth of mapping was exported from CLC and coverage analysis was performed with R script.

We then performed a blind analysis in the setting without prior knowledge of the virus subtype in the mixed virus sample. We first assembled all of the filtered reads de novo, then extracted and blasted the contigs with a length >500 bp in the NCBI nt database. Contigs aligned to the HA and NA was selected to determine the virus subtype in the mixture.

For clinical samples, all the contigs from de novo assembly were extracted and blasted in the NCBI nt database. For each sample, the genome of the best bit score in the database was selected as a reference for reads mapping. Consensus sequences were then extracted from the mapping to acquire the virus genome.

**RESULTS**

**Sensitivity of M-RTPCR for WGS of Single-subtype Influenza A Virus**

For WGS of single subtype influenza A virus, we normalized the reads number to 500,000 for each dilution. In the dilutions of 625 and 125 copies/reaction (125 and 25 copies/μL), M-RTPCR could get the virus complete genome with high depth (Table 1). In the lowest dilution (25 copies/reaction), M-RTPCR had 6 segments fully covered, including HA and NA, from which the virus subtype could be determined. The three dilutions produced close total reads mapped, although the virus titer differed 25-fold between the highest and lowest dilutions.

**Sensitivity to Differentiate Mixed Viruses**

To assess the ability of M-RTPCR to differentiate mixed subtype virus with biased copies, we manually mixed a certain copies of PR8 genome with 100-, 200-, 500-, and 1000-fold higher copies of G1 genome to construct four subtype-mixed samples.

**Table 1. Number of ReadsMapped to Each Segment of PR8 in Total of 500,000 Reads Obtained from Different PR8 Dilution (Copies/Reaction)**

<table>
<thead>
<tr>
<th>Segment</th>
<th>625$^d$</th>
<th>125$^d$</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB2</td>
<td>37,686</td>
<td>32,278</td>
<td>8,841</td>
</tr>
<tr>
<td>PB1</td>
<td>45,383</td>
<td>52,697</td>
<td>64</td>
</tr>
<tr>
<td>PA</td>
<td>43,144</td>
<td>42,802</td>
<td>44,316</td>
</tr>
<tr>
<td>HA</td>
<td>53,614</td>
<td>37,828</td>
<td>46,815</td>
</tr>
<tr>
<td>NP</td>
<td>39,044</td>
<td>15,428</td>
<td>7</td>
</tr>
<tr>
<td>NA</td>
<td>54,844</td>
<td>38,626</td>
<td>74,331</td>
</tr>
<tr>
<td>MP</td>
<td>106,835</td>
<td>138,842</td>
<td>236,297</td>
</tr>
<tr>
<td>NS</td>
<td>111,143</td>
<td>129,648</td>
<td>79,927</td>
</tr>
<tr>
<td>Total mapped</td>
<td>491,693</td>
<td>488,149</td>
<td>490,527</td>
</tr>
</tbody>
</table>

**Note.** $^*$Segments were partially covered or uncovered. Segment was completely covered if no label assigned. $^d$Whole genome was covered in this dilution.
After M-RTPCR and PGM sequencing, 390,000 reads were sampled for each mixture. We found the predominant virus G1 could get complete genome covered in all the four mixture samples. Thus, the following evaluation was focused on genome coverage of the minor virus PR8. We obtained a nearly complete genome of PR8, with only 155 and 92 bp unmapped in PA and PB2, respectively, in the copy-rate 100 mixture (Figure 1). Six segments were ≥90% covered in the copy-rate 200 mixture, including HA and NA (Figure 1) (Table S1, www.besjournal.com for details). Fewer segments were fully mapped, including HA and NA, in the following mixtures.

De Novo Assembly and BLASTn Analysis of Influenza A Reads in Mixed Subtype Virus Sequencing

In the blind analysis, contigs with length >500 bp were aligned to the NCBI nt database to identify potential subtypes existing in the mixture samples. As expected, the predominant virus G1 assembled long contigs aligned to H9 and N2, in which the longest contig covered >80% of the region in the targeted H9 and N2 (1299-1740 bp for H9 and 1435-1477 for N2). For the minor virus PR8 (H1N1), both H1 and N1 matching contig was detected in the copy-rate 100, 200, and 500, in which the longest H1 contig ranged from 850-1799 bp and N1 contig from 1300-1401 bp in the 3 mixture samples. In the highest copy-rate mixture, the H1 matching contig was missing and the N1 matching contig was only 674 bp.

Clinical Sample Sequencing

Four clinical samples tested as influenza A virus positive were subjected to M-RTPCR. Gel electrophoresis showed two samples (Figure S1 www.besjournal.com for details, Swab -2 and 3) obtained a good amplification with clear cDNA bands, while another two samples (Swab 1 and 4) achieved poor amplification. The two samples with good amplification produced sufficient cDNA products were sequenced on the PGM platform. For another two samples with less cDNA amplified, 1 ng cDNA was subjected to library construction and sequencing on the Illumina MiSeq platform. From the blast results, the genome of A/Singapore/H2013.422c/2013 (H3N2) was most similar to the contigs assembled from the four samples. Thus, the eight segments of this virus were selected as reference for reads mapping. As expected, the good amplified samples, the swab-2 and 3, obtained complete genome mapping with high depth (Table 2). Moreover, we could directly assemble long contigs covering at least 80% of the region for segments of the virus in these two samples. Virus quantification indicated that the virus titers in these two samples were 296 copies/μL and 320 copies/μL, respectively. For the two poorly amplified samples, swab 1 only got MP and NS fully covered, and HA and NA in swab 4. The virus titers in these two samples were 18 and 22.8 copies/μL, respectively, which were below the titer threshold of the M-RTPCR method we determined above.

Table 2. Number of Reads Mapped to Each Segments of the H3N2 Reference Genome in Four Clinical Samples

<table>
<thead>
<tr>
<th>Segment</th>
<th>Swab-1</th>
<th>Swab-2</th>
<th>Swab-3</th>
<th>Swab-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB2</td>
<td>5</td>
<td>47,024</td>
<td>36,304</td>
<td>357</td>
</tr>
<tr>
<td>PB1</td>
<td>0</td>
<td>37,532</td>
<td>27,641</td>
<td>18</td>
</tr>
<tr>
<td>PA</td>
<td>28,450</td>
<td>55,865</td>
<td>55,826</td>
<td>12</td>
</tr>
<tr>
<td>HA</td>
<td>32</td>
<td>50,830</td>
<td>58,459</td>
<td>5,417</td>
</tr>
<tr>
<td>NP</td>
<td>0</td>
<td>46,246</td>
<td>43,074</td>
<td>6</td>
</tr>
<tr>
<td>NA</td>
<td>2</td>
<td>68,710</td>
<td>54,493</td>
<td>52,499</td>
</tr>
<tr>
<td>MP</td>
<td>19,556</td>
<td>137,516</td>
<td>179,006</td>
<td>24</td>
</tr>
<tr>
<td>NS</td>
<td>35,490</td>
<td>43,873</td>
<td>40,854</td>
<td>0</td>
</tr>
<tr>
<td>Total mapped</td>
<td>83,535</td>
<td>487,614</td>
<td>495,657</td>
<td>58,333</td>
</tr>
<tr>
<td>Total obtained</td>
<td>253,676</td>
<td>525,319</td>
<td>569,567</td>
<td>542,082</td>
</tr>
</tbody>
</table>

Note. *Samples with whole virus genome covered, or segments were completely covered. †Total reads obtained from each clinical swab.
DISCUSSION

As a worldwide spreading pathogen, influenza A virus has caused several pandemics in the last century and has received significant attention within the disease control field[9]. The advent of NGS technology has greatly facilitated large-scale WGS of influenza virus swiftly and economically. The M-RTPCR method was developed for high-throughput sequencing and/or reverse genetics of influenza A virus, which could simultaneously amplify eight genomic RNA segments in a single reaction, independent of subtypes. This method applied two primers accounting for the U/C variation at position 4 of the 3' terminus and modestly improved the amplification efficiency. We used M-RTPCR for influenza A virus genome amplification in the NGS for its convenience and economy. A systematic evaluation of this method set a reference for other applications of NGS for influenza A virus WGS.

**Figure 1.** Coverage map of the PR8 influenza genome mixed with 100 to 1000-fold titer of predominant subtype G1. The y axis indicates the coverage depth of a given nucleotide in the reference. The x axis corresponds to nucleotide positions in the reference. Most of the eight segments obtained decent coverage in the copy-rate 100 and 200 mixture samples (G1 to PR8). The uncovered region had no line assigned above. CR: Copy-rate of G1 to PR8.
Our data showed that M-RTPCR could achieve a complete genome with a virus titer of 25 copies/μL when one subtype existed, which corresponded to a cycle threshold value of 33.7 in the TaqMan RT-PCR assay. Moreover, influenza virus could still be subtyped when the genome titer decreased to 5 copies/μL because HA and NA could be completely covered. Furthermore, there were challenges to sequence low-titer influenza virus mixed with high-titer ones because of exponential amplification in the RT-PCR step, so we mixed 125 copies/μL of the minor virus PR8, with a different titer of the predominant virus G1, in the assay of subtyping mixed virus. A nearly complete genome of the minor virus was obtained when the predominant virus had a titer higher than 100 folds (Figure 1). The minor virus subtype could still be determined when the major virus titer increased to 200 folds. Moreover, using de novo assembly, we detected the HA and NA of the minor virus in further biased copy-rate 500, with which the subtype of minor virus could be determined in the mixture.

We further applied the M-RTPCR method for direct WGS of four clinical samples with a low titer of virus (CT value of 31-34). As expected, two samples with a virus titer above the method threshold (25 copies/μL) achieved complete genome with high depth, while samples with an under-threshold titer of virus only had a partial genome covered. We also sequenced the good amplification samples on the MiSeq platform, and obtained similar results as PGM, indicating that application of M-RTPCR was universal at the two most popular benchtop NGS platforms. Since virus isolation is prone to introduce site mutations in influenza A virus\textsuperscript{10}, direct WGS from the clinical samples is critical to characterize virus genotype at native profile.

Finally, it should be underscored that these results were obtained by directly diluting virus genome to different titers, not by extracting RNA from virus of different copies. The reason we applied this strategy was that we aimed to evaluate the sensitivity of the M-RTPCR for WGS of influenza A virus, exempting the bias existing in RNA extraction with different methods or kits. Optimizing the RNA extraction procedure and better RNA preserving status is favorable for WGS of influenza virus. Moreover, we used the PR8 (H1N1) virus to evaluate the sensitivity of the M-RTPCR method in this study; however, this method also demonstrated good amplification for other influenza A subtypes, both in the Zhou et al. study\textsuperscript{13} and our routine surveillance program\textsuperscript{11-12}. Therefore, the sensitivity of M-RTPCR we detected herein is also applicable to WGS of other influenza A virus subtypes.

**CONCLUSION**

Our study provided evaluation of a single-reaction amplification method for WGS of influenza A virus using NGS. The M-RTPCR was a robust and convenient amplification method which could obtain whole genome of influenza A virus at low titers. Moreover, this method could differentiate mixed-virus even the two virus titers differ by as much as 200 folds. Using de novo assembling and blasting, the method could get hints of minor virus at larger copy disparity. Our data set reference for others using the NGS platform to WGS of influenza A virus.

Received: August 18, 2015; Accepted: January 7, 2016

**REFERENCES**