Original Article

An Improved Barcoded Oligonucleotide Primers-based Next-generation Sequencing Approach for Direct Identification of Viral Pathogens in Clinical Specimens*

WANG Chun Hua[1], NIE Kai[1], ZHANG Yi[1], WANG Ji[1], ZHOU Shuai Feng[1,2], LI Xin Na[1], ZHOU Hang Yu[1], QI Shun Xiang[3,#], and MA Xue Jun[1,#]

1. Key Laboratory for Medical Virology, National Health and Family Planning Commission, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China; 2. Center for Disease Prevention and Control of Hunan Province, Changsha 410005, Hunan, China; 3. Institute for Viral Disease Control and Prevention, Center for Disease Control and Prevention of Hebei, Shijiazhuang 050000, Hebei, China

Abstract

Objective To provide a feasible and cost-effective next-generation sequencing (NGS) method for accurate identification of viral pathogens in clinical specimens, because enormous limitations impede the clinical use of common NGS, such as high cost, complicated procedures, tremendous data analysis, and high background noise in clinical samples.

Methods Viruses from cell culture materials or clinical specimens were identified following an improved NGS procedure: reduction of background noise by sample preprocessing, viral enrichment by barcoded oligonucleotide (random hexamer or non-ribosomal hexanucleotide) primer-based amplification, fragmentation-free library construction and sequencing of one-tube mixtures, as well as rapid data analysis using an in-house pipeline.

Results NGS data demonstrated that both barcoded primer sets were useful to simultaneously capture multiple viral pathogens in cell culture materials or clinical specimens and verified that hexanucleotide primers captured as many viral sequences as hexamers did. Moreover, direct testing of clinical specimens using this improved hexanucleotide primer-based NGS approach provided further detailed genotypes of enteroviruses causing hand, foot, and mouth disease (HFMD) and identified other potential viruses or differentiated misdiagnosis events.

Conclusion The improved barcoded oligonucleotide primer-based NGS approach is simplified, time saving, cost effective, and appropriate for direct identification of viral pathogens in clinical practice.

Key words: NGS; Barcoded oligonucleotide primers; Virus identification; Clinical specimen

Biomed Environ Sci, 2017; 30(1): 22-34 *doi: 10.3967/bes2017.003* *ISSN: 0895-3988*

www.besjournal.com (full text) *CN: 11-2816/Q*

*This work was supported by the China Mega-Project for Infectious Disease (2016ZX10004-101, 2016ZX10004-215); Beijing Municipal Science & Technology Commission Project (D151100002115003); and Guangzhou Municipal Science & Technology Commission Project (2015B2150820).

#Correspondence should be addressed to MA Xue Jun, Professor, PhD, Tel/Fax: 86-10-58900810, E-mail: maxj@ivdc.chinacdc.cn; QI Shun Xiang, Senior technologist, MS, E-mail: hbcdc999@126.com

Biographical note of the first author: WANG Chun Hua, female, born in 1988, Doctoral degree candidate, majoring in pathogen biology.

INTRODUCTION

Viruses are a cause of serious health problems in humans, both in developed and developing countries; in particular, some pathogenic viruses have a potential for rapid and global spread with high morbidity and mortality. The last 30 years of the 21st century have witnessed some outbreaks of significant public health issues caused by viruses, such as Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) in 2003[1-3], swine-origin H1N1 influenza virus in 2009[4-5], avian-origin influenza A H7N9 virus in 2013[6-8], Middle East Respiratory Syndrome (MERS-CoV) in 2012[9-12], and Ebola virus from 2014 to 2015[13-14]. Therefore, accurate identification and analysis of the causative agent of viral infectious disease is extremely urgent and crucial for active control and prevention of virus spread.

In recent years, next-generation sequencing (NGS) has been an emerging method for the identification of viral pathogens. It does not require a priori knowledge of potential infectious agents and is a more powerful tool than traditional viral diagnostic methods, including electron microscopy, cell culture, antigen detection, nucleic acid detection, cytology, histology, and serology[15]. Multiple viruses can be detected simultaneously, and novel or highly divergent viruses can be discovered and genetically characterized using NGS[16-24]. However, on account of the low virus abundance and high background noise in clinical specimens, the clinical application of NGS in virus detection is still limited. Moreover, other limitations, such as the relatively high cost of library preparation and sequencing, complicated and tedious procedures, as well as tremendous data processing and analysis, also hinder the widespread use of NGS in clinical practice.

At present, there are many ways to increase the signal-to-noise ratio (ratio of viral genome to host genome), such as ultracentrifugation, filtration, nuclease treatment, sequence-independent amplification (like random PCR), or a combination of these[25]. Particularly, a combination method of centrifugation, filtration, nuclease treatment, and random-PCR (rPCR) showed the greatest increase in the proportion of viral sequences[26]. In addition, rPCR is a widely used and efficient method in which reverse transcription is performed by using oligonucleotides made up of random hexamers tagged with a known sequence for the enrichment of low-abundance viral nucleic acids[27-30]. Recently, a substracted primer set named non-ribosomal hexanucleotides (hexanucleotides) was designed to further reduce noise for reverse transcription, and its viral enrichment ability was verified by cDNA representational difference analysis (RDA) and compared with that of random hexamers[31]. NGS-based evaluation of the viral enrichment abilities of both random hexamers and hexanucleotides is yet unreported. Directly pooling samples at the very beginning (without barcodes) or pooling individual libraries (with barcoded adaptors) for NGS may reduce sequencing costs, but the subsequent verification tests of the interested sample from pooled samples or the individual library construction of each pooled sample are rather laborious and time consuming. In addition, fragmented DNA or RNA libraries are always recommended to acquire much information from the original sample; however, the shear time is difficult to control because DNA or RNA originates from samples with different virus abundance and this treatment may increase the occurrence of artefactual recombination[32]. Moreover, the processing and analysis of tremendous NGS data is always complex, time consuming, and highly dependent on excellent professional bioinformaticians, which is impractical in many laboratories.

To address these limitations, in the present study, we describe a simplified, rapid, and relatively cost-effective NGS approach in which the barcoded random hexamer or non-ribosomal hexanucleotide primers are used in the reverse transcription step for viral enrichment followed by sample pooling of each enriched product in one tube as a single sample for fragmentation-free library construction and sequencing, and the high-throughput sequencing data are rapidly analyzed by an in-house pipeline. We demonstrate that multiple viruses can be accurately identified from cell culture materials or clinical specimens and verify that the viral enrichment abilities of both primer sets are nearly comparative in model experiments (sequencing three representative viruses). Furthermore, direct testing of clinical specimens associated with hand, foot, and mouth disease (HFMD) using this improved hexanucleotide primer-based NGS approach provides further detailed genotypes of viral pathogens and identifies other potential viruses or differentiated misdiagnosis events appearing in the clinic.