Journal of Virological Methods 199 (2014) 68–75

Contents lists available at ScienceDirect

Journal of Virological Methods

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

Deep sequencing approach for genetic stability evaluation of influenza A viruses

Bella Bidzhieva, Tatiana Zagorodnyaya, Konstantinos Karagiannis, Vahan Simonyan, Majid Laassri*, Konstantin Chumakov

Center for Biologics Evaluation and Research, US Food and Drug Administration, 1401 Rockville Pike, HFM 470, Rockville 20852, MD, USA

Article history:
Received 10 September 2013
Received in revised form 9 December 2013
Accepted 17 December 2013
Available online 7 January 2014

Keywords:
Sequence heterogeneity
Profiles of mutations
Sequencing libraries preparation
Pyrosequencing
Illumina Solexa sequencing

A B S T R A C T

Assessment of genetic stability of viruses could be used to monitor manufacturing process of both live and inactivated viral vaccines. Until recently such studies were limited by the difficulty of detecting and quantifying mutations in heterogeneous viral populations. High-throughput sequencing technologies (deep sequencing) can generate massive amounts of genetic information and could be used to reveal and quantify mutations. Comparison of different approaches for deep sequencing of the complete influenza A genome was performed to determine the best way to detect and quantify mutants in attenuated influenza reassortant strain A/Brisbane/59/2007 (H1N1) and its passages in different cell substrates. Full-length amplicons of influenza A virus segments as well as multiple overlapping amplicons covering the entire viral genome were subjected to several ways of DNA library preparation followed by deep sequencing using Solexa (Illumina) and pyrosequencing (454 Life Science) technologies. Sequencing coverage (the number of times each nucleotide was determined) of mutational profiles generated after 454-pyrosequencing of individually synthesized overlapping amplicons were relatively low and insufficiently uniform. Amplification of the entire genome of influenza virus followed by its enzymatic fragmentation, library construction, and Illumina sequencing resulted in high and uniform sequencing coverage enabling sensitive quantitation of mutations. A new bioinformatic procedure was developed to improve the post-alignment quality control for deep-sequencing data analysis.

Published by Elsevier B.V.

1. Introduction

Genetic variability and plasticity of genomes are inherent properties of RNA viruses and have many profound implications for their replication, evolution, and pathogenesis (Domingo et al., 1985; Hansen et al., 2004). Many important virus properties cannot be explained by knowing a mere consensus sequence, but require quantitation of all mutants that are present in viral stocks. Such properties include virulence, which is critical for vaccine development and manufacture, and antigenicity that may also be affected by mutations emerging in antigenic sites. Therefore assessment of genetic stability is an important part of pre-licensure evaluation and quality control of live and inactivated viral vaccines. Development of assays for discovery of mutations in the entire viral genome as well as their sensitive quantitation is a challenge. Conventional sequencing approaches are suitable for discovery of mutations that are present in substantial amounts, usually more than 20%, and sequencing of large sets of virus clones (plaques) is labor-intensive and slow.

There are also indirect approaches based on analysis of electrophoretic mobility in gels (Orita et al., 1989), which are insufficiently sensitive and do not allow mutations to be located accurately. Matrix-assisted laser desorption/ionisation time of flight (MALDI-TOF) mass spectrometry (Amexis et al., 2001) and hybridization with microarrays of short oligonucleotides (Cherkasova et al., 2003; Laassri et al., 2012) are more sensitive, but are relatively laborious and may require follow-up by direct sequencing. A highly sensitive mutant analysis by PCR and restriction enzyme cleavage (MAPREC) (Chumakov, 1999; Chumakov et al., 1991) can detect and quantify mutants at levels as low as 0.1% of the viral population, but is only suitable for analysis of one mutation in each assay.

Next generation of sequencing technologies that are also referred to as massively parallel sequencing (MPS) or deep sequencing can produce millions of sequencing reads in one run. They are used for de novo sequencing of large genomes, metagenomics studies, screening for genomic markers, analysis of a transcriptome, and many other applications (Bainbridge et al., 2006; Cheval et al., 2011; Greninger et al., 2010; Kuroda et al., 2010; Nakamura et al., 2009; Pettersson et al., 2008; Satkoski et al., 2008; Torres et al., 2008; Wheeler et al., 2008). Recently we have demonstrated that MPS can also be used to monitor genetic stability of oral polio vaccine

* Corresponding author. Tel.: +301 827 2636; fax: +301 827 4622.
E-mail address: majid.laassri@fda.hhs.gov (M. Laassri).

0166-0934/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.jviromet.2013.12.018

(Neverov and Chumakov, 2010) and could replace the currently used MAPREC assay.

Influenza A viruses are enveloped, single stranded RNA viruses belonging to the orthomyxoviridae family (Lamb and Krug, 2007) that also contains four other viral species: influenza B virus, influenza C virus, thogotovirus, and isavirus. The genome of influenza A virus is about 13.6 kb in size and consists of 8 negative-polarity RNA segments that encode 10 or 11 proteins. Its genome is highly variable due to low fidelity of RNA polymerase and reassortment (segment exchange) between co-infecting strains (Steinhauer et al., 1992). New virus mutants emerge continuously allowing viruses to evade host immunity and cause recurring annual epidemics and occasionally pandemics. Because of this genetic and antigenic plasticity, influenza vaccines must be frequently reformulated to include antigens of the currently circulating strains. Both live and inactivated influenza vaccines are produced by reassortment with high-growth strains to enable adequate yields in eggs or cell substrates used for vaccine production (Girard et al., 2006; McCarthy and Kockler, 2004). Adaptation to growth in different cells can lead to changes in viral receptor-binding region, and also in protective epitopes, because they overlap each other. Therefore it may be desirable to monitor genetic stability of viruses used in vaccine manufacture to ensure that their antigenic structure remains unchanged. Here we explore the utility of deep sequencing methods for monitoring the consistency of influenza A vaccines. We also describe optimized protocols for full-length amplification of influenza A genomes and new bioinformatic tools to analyze the data and to identify artifacts generated during PCR amplification and sequencing procedures. The detailed dynamics of mutant accumulations and the biologic effects will be published elsewhere.

2. Materials and methods

2.1. Amplification of the entire genome of influenza A viruses

A/Brisbane/59/2007 (H1N1), A/Uruguay/716/2007 (H3N2), and A/California/07/2009 (H1N1) viruses were kindly provided by Dr. Zhiping Ye (Division of Viral Products, CBER/FDA). A/Ohio/02/2007(H1N1) and A/swine/Italy/50175/2007 (H1N1) viruses were provided by Dr. Mikhail Matrosovich (Philipps University, Germany). Viral RNA was isolated from 140 μl of virus-containing cell culture fluids using QIAamp Viral RNA Mini Kit (Qiagen, Valencia, CA) according to the manufacturer's protocol. The RNA was eluted in a final volume of 60 μl of sterile RNase-free water. cDNA was synthesized using reverse transcriptase SuperScript III (SSIII, Invitrogen, Carlsbad, CA) with universal forward primer Universal_F (Table 1). cDNA synthesis was performed in a reaction containing 10 μl of isolated viral RNA, 1.6 μM of forward universal primer, 600 units of SSIII, 0.5 mM dNTPs, 50 mM DTT and 1× first strand buffer in a total volume of 50 μl. The reaction mixture was incubated at 55 °C for 2 h followed by enzyme inactivation at 70 °C for 15 min. Four microliters of viral cDNA was used as a template for PCR amplification in a total reaction volume of 50 μl containing 0.2 μM of each universal primer, 200 μM of dNTP, 3 mM $MgCl_2$, and 1 unit of Fusion DNA polymerase (New England BioLabs, Ipswich, MA). A pair of universal PCR primers was complementary to the 3′ and 5′ ends that are identical for all RNA segments. Forward and reverse primers also contained sequences of T7 and SP6 promoter, respectively (Table 1). PCR cycling conditions were as follows: an initial denaturation at 98 °C for 30 s followed by 35 cycles of 98 °C for 10 s, 55 °C for 30 s, 72 °C for 5 min, and a final elongation at 72 °C for 10 min. The resulting 8 amplicons representing all segments of viral genome were separated by electrophoresis in 1% agarose gels with ethidium bromide (Lonza, Rockland, ME) and visualized using Kodak Gel Logic 200 Imaging System and Kodak Molecular Imaging Software (Carestrem Health, Inc. Rochester, NY).

2.2. Preparation of DNA samples for 454-pyrosequencing

Viral cDNA samples were prepared from the initial stock of A/Brisbane/59/2007 (H1N1) strain, its first and tenth passages in embryonated chicken eggs, as well as MDCK and DF-1 cell cultures. Then PCR was performed using the specific primers (Table 1) to obtain 500–800 bp overlapping amplicons that covered the entire genome of A/Brisbane/59/2007 (H1N1) reassortant virus. All eight segments sequences of initial reassortant virus were submitted to GenBank under the following access numbers: JX414009-JX414016. These DNA sequences were also used as a reference genome for further deep sequencing data analysis. The amplification procedure was performed in a total volume of 50 μl contained 0.2 μM of each specific primer, 200 μM of dNTP, 3 mM $MgCl_2$, and 5 Unit Hot Start Taq Plus DNA polymerase (Qiagen, Valencia, CA). PCR cycling conditions were as follows: an initial denaturation for 15 min at 95 °C followed by 35 cycles of 95 °C for 30 s, 54 °C for 30 s and 72 °C for 1 min, and a final elongation at 72 °C for 7 min. All PCR amplicons were purified with QIAGEN PCR purification kit (Qiagen, Valencia, CA), then mixed together at an equimolar ratio. Then, DNA libraries suitable for pyrosequencing analysis were prepared and subjected to 454-pyrosequencing at the Laboratory of Molecular Technology, Advanced Technology Program, SAIC (Frederick, MD) or Beckman Coulter Genomics (Beverly, MA). The sequencing data were analyzed by custom software developed in our laboratory.

2.3. Preparation of samples for Illumina/Solexa sequencing

The last passages of A/Brisbane/59/2007 (H1N1) reassortant viruses in MDCK and DF-1 cell cultures were used to prepare DNA libraries suitable for Illumina sequencing. The DNA library preparation procedure included full genome amplification with universal forward and reverse primers (Table 1) as described above. The PCR product was purified by QIAGEN-PCR purification kit (Qiagen, Valencia, CA) and fragmented by NEBNext dsDNA Fragmentase (New England BioLabs, Ipswich, MA) according to the manufacture protocols. Each sample was incubated with the enzyme for 15 min, 20 min, and 30 min to obtain the optimal fragment sizes 350–400 bp suitable for Illumina/Solexa sequencing.

The DNA fragments were analyzed for DNA size and quality on BioAnalyzer using a high sensitivity kit (Agilent Technologies, Inc., Santa Clara, CA). Then, the DNA fragments were used for library preparation with NEBNext® DNA Sample Prep Reagent Set 1 (New England BioLabs, Ipswich, MA). The primers containing multiplexing index adapters ("barcodes") to identify sequencing reads belonging to each sample were synthesized according to the specifications provided by Illumina Inc. for multiplex pair-end sequencing. Deep sequencing was performed at Macrogen (Seoul, Korea) using HiSeq2000 (Illumina) producing 101 bp paired-end reads.

2.4. Algorithms and custom software for sequencing data analysis

The prototype DNA sequence of A/Brisbane/59/2007 (H1N1) reassortant virus was used as a template for alignment of individual sequencing reads. First, sequencing reads with low quality (phred) score were removed from the data set, and the remaining sequences aligned with reference influenza virus sequence using custom software implemented in highly integrated virtual environment (HIVE) computer cluster that will be described in details elsewhere.