Whole-Genome Sequencing During Measles Outbreaks

Paul A. Rota and Bettina Bankamp

Centers for Disease Control and Prevention, Atlanta, Georgia

(See the brief report by Gardy et al on pages 1574–8.)

Genetic characterization of wild-type measles viruses is an important component of laboratory-based surveillance for measles virus. The genetic information can classify measles viruses as endemic or imported and can help map the transmission pathways of the virus. Since virologic evidence demonstrating the absence of an endemic genotype of measles virus is one of the key criteria for verification of measles elimination [1], adequate virologic surveillance must be conducted on a global scale. The current recommendation is to obtain appropriate samples for viral genotyping from at least 80% of outbreaks [2]. In addition, sequence analysis is currently the only method available to confirm measles vaccine–associated rash and fever. Distinguishing between reactions to vaccine and cases caused by wild-type virus is an important part of laboratory support for outbreak investigations [3].

Virologic surveillance is performed by the World Health Organization’s (WHO’s) Global Measles and Rubella Laboratory Network (LabNet). All of the LabNet laboratories use a standard protocol for determining the genotype of measles virus. The WHO-recommended protocol consists of analyzing the sequence of the 450 nucleotides coding for the carboxy-terminal 150 amino acids of the nucleoprotein (N-450). Sequence data are compared to a set of WHO reference sequences representing each of the 24 recognized genotypes [3–5]. Sequence data are submitted to the WHO Global Measles Sequence Database, which currently contains >20,000 sequences. In addition, sequences may be assigned to a specific named lineage [4]. Named lineages represent at least 50 identical N-450 sequences reported within the previous 12 months and are named on the basis of the earliest virus detected in the lineage. Despite N-450 being one of the most variable regions in the measles genome, named lineages can contain several hundred or more strains with identical N-450 sequences.

The measles virus genotyping protocols were initially developed in 1998. Although generating N-450 sequences was essential for determining the overall genetic variability of measles viruses and for detecting new genotypes, the use of this sequencing “window” has limitations. Because named lineages can circulate throughout the world for several years, analysis of N-450 sequences is often not sufficient to distinguish between continued circulation of viruses in a single chain of transmission or multiple importations of viruses from various sources. Since this distinction is vital for verification of measles elimination, the LabNet began to study increasing the size of the “sequencing window” to improve the resolution of the sequence data. In addition to N-450, laboratories compared the sequences of the full-length phosphoprotein and hemagglutinin genes, as well as the sequence of an untranslated region that lies between the coding regions for the matrix and fusion proteins. The results showed that viruses with identical N-450 sequences were often representatives of different lineages [6, 7]. Of course, increasing the size of the sequencing target increases the complexity of the sequencing protocols and puts greater demand on the LabNet laboratories. Ultimately, the optimum sequencing window would be the sequence of the complete viral genome. Improvements in sequencing technology now make whole-genome sequencing (WGS) feasible for studying the molecular epidemiology of measles virus.

There are several different approaches to WGS for measles virus. Amplification of the genome in several large fragments with virus-specific primers followed by Sanger sequencing of the amplicons, as described by Gardy et al, has proven successful for years and, to our knowledge, produced all of the complete genomes currently available on GenBank. The methods are well established, equipment is widely available, and the procedure is relatively affordable. High-throughput sequencing of such amplicons with Illumina or 454 next-generation sequencing technologies (NGS) has been used to generate WGS for other RNA viruses, including norovirus, enterovirus A-71, hepatitis A virus, and hepatitis C virus [8–11]. An NGS approach with production of libraries...
based on random primers is also proving useful for obtaining WGS data for measles virus (Bankamp and Rota, unpublished data). Advantages of NGS are the speed of data generation, especially when multiplexing samples, and the avoidance of the use of gene-specific primers. Such methods are under development in several laboratories in the LabNet.

The article by Gardy et al in this issue of The Journal of Infectious Diseases describes the investigation of a measles outbreak that occurred in Vancouver, Canada, following the XXI Olympic Winter Games. Following standard genetic analysis of viral samples, which detected imported genotypes H1 and D8, the authors showed that it is technically feasible to quickly sequence the genomes of numerous measles virus isolates. This article represents the largest measles virus genomics project to date, collecting WGS from 27 viral isolates. Based on the information obtained from these genomes, a polymerase chain reaction assay to detect single-nucleotide variants (SNV) was developed to type clinical specimens. While the SNVs were outbreak specific, a similar approach combining WGS with SNV analysis could be designed to investigate other large or long-lasting outbreaks. This report is also the first to use WGS data for molecular epidemiology analysis of a measles outbreak. The increased resolution provided by WGS data was useful for analyzing pathways of transmission. The data showed that that there were no nucleotide substitutions in isolates that were recovered from person-to-person transmission and established that a single introduction of genotype H1 virus led to the expansion of the outbreak beyond Vancouver. The observed increase in genetic diversity as the outbreak progressed and the tracing of one measles virus variant along a major highway are additional examples of the power of analyzing complete genomic sequences.

Routine use of WGS of measles viruses for the purpose of molecular epidemiology will bring several challenges. Use of NGS requires agreement among participating laboratories to develop adequate criteria for quality control for the various sequencing platforms and analysis software. Ideally, a standardized analysis work flow for participating LabNet laboratories could be developed. Although all approaches to WGS described above can be applied to virus isolates, WGS of patient samples with varying viral loads will often provide incomplete coverage of the genome. Therefore, guidelines need to be developed to ensure that WGS provides high-quality data that can be used for molecular epidemiologic studies. The high cost in both equipment and reagents makes WGS less feasible than standard genotyping for most LabNet laboratories; therefore, laboratories will need to collaborate by sharing sequences and samples.

Notes

Disclaimer. The commentary in this report is that of the authors and does not necessarily represent the official position of the Centers for Disease Control and Prevention.

Potential conflict of interest. Both authors: No reported conflicts.

Both authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References