AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Spectrum

| Open Peer Review | Computational Biology | Research Article

Cost-effective complete genome sequencing using the MinION platform for identification of recombinant enteroviruses

Yeh-Sheng Chien,[1] Feng-Jui Chen,[1,2] Han-Chieh Wu,[1] Chieh-Hua Lin,[3] Wen-Chiung Chang,[1] David Perera,[4] Jyh-Yuan Yang,[5] Min-Shi Lee,[1] Yu-Chieh Liao[3]

AUTHOR AFFILIATIONS See affiliation list on p. 13.

ABSTRACT Enteroviruses (EVs) are a group of viruses that cause various human illnesses. While the CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) method can generate VP1 gene fragments for enterovirus genotyping, it is unable to detect recombinant strains. Recent advances in viral genome sequencing using next-generation sequencing technologies have enabled comprehensive analyses. However, the high cost poses a challenge for widespread adoption. To address this issue, this study proposes a cost-effective approach for generating complete enterovirus genome sequences using the Oxford Nanopore MinION sequencer. This protocol not only facilitates the generation of accurate genome sequences for various enterovirus strains but also allows for the differentiation of co-infections from viral isolates. In addition, the method can generate polyprotein sequences as well as peptide sequences of the upstream ORF (uORF) whose expression can impact virus infection. Through the analysis of complete enterovirus genomes, this study successfully identified seven enterovirus A71 isolates obtained during the 2018 enterovirus outbreak in Malaysia and Taiwan as recombinants between enterovirus A71 and coxsackievirus A2. Furthermore, our study has made a significant discovery by establishing a strong correlation between uORF trees and the epidemics of EVA71. This finding highlights the potential of uORF sequences as valuable indicators for monitoring and understanding the spread of EVA71 infections. We also identified notable amino acid changes in the transmembrane domain of the uORF protein within a newly identified lineage. These findings provide crucial insights into the molecular characteristics and evolutionary dynamics of EVA71, offering valuable information for future research and intervention strategies.

IMPORTANCE By employing a cost-effective approach for complete genome sequencing, the study has enabled the identification of novel enterovirus strains and shed light on the genetic exchange events during outbreaks. The success rate of genome sequencing and the scalability of the protocol demonstrate its practical utility for routine enterovirus surveillance. Moreover, the study's findings of recombinant strains of EVA71 and CVA2 contributing to epidemics in Malaysia and Taiwan emphasize the need for accurate detection and characterization of enteroviruses. The investigation of the whole genome and upstream ORF sequences has provided insights into the evolution and spread of enterovirus subgenogroups. These findings have important implications for the prevention, control, and surveillance of enteroviruses, ultimately contributing to the understanding and management of enterovirus-related illnesses.

KEYWORDS enterovirus, recombinant, next-generation sequencing, nanopore, upstream ORF

Enteroviruses (EVs) belong to the *Picornaviridae* family and are widespread and environmentally stable RNA viruses (1). They are characterized by a conserved

Editor Day-Yu Chao, National Chung Hsing University, Taichung, Taiwan

Ad Hoc Peer Reviewer Michael Owusu, Kumasi Centre for Collaborative Research in Tropical Medicine (KCCR), Kumasi, Ghana

Address correspondence to Yu-Chieh Liao, jade@nhri.edu.tw, or Min-Shi Lee, minshi.lee@hotmail.com.

Yeh-Sheng Chien and Feng-Jui Chen contributed equally to this article. The author order was determined based on seniority.

The authors declare no conflict of interest.

See the funding table on p. 13.

Received 14 June 2023
Accepted 30 August 2023
Published 13 October 2023

genomic structure, consisting of a single-stranded RNA genome ranging in size from 7.2 to 8.5 kb (2). The capsid proteins, VP1 to VP4, are encoded in the P1 region of the enterovirus genome, whereas the nonstructural proteins are encoded in the remaining P2 and P3 regions (3). The VP1 capsid gene sequencing has shown a strong correlation with serotypes determined through virus neutralization tests, making it an excellent tool for genotyping enteroviruses. To facilitate the sequencing of VP1 sequences, a consensus degenerate hybrid oligonucleotide primer (CODEHOP) approach was proposed (4, 5). In 2019, the existence of an upstream open reading frame (uORF) that encodes an additional polypeptide called the uORF protein (UP) was proposed and demonstrated to play a role in modulating gut infection (1). However, research on UP is very limited, and its evolution remains unclear. The *Enterovirus* genus encompasses 15 species, namely *Enterovirus A–D* (human enteroviruses), *Enterovirus E–L* and *Rhinovirus A–C*, as defined by the International Committee on Taxonomy of Viruses (ICTV) (6). These species contain over 200 serotypes, which are associated with a range of diseases. For example, Coxsackievirus A16 and Enterovirus A71 are known to cause hand-foot-mouth disease (7), while Polioviruses 1–3 are responsible for poliomyelitis (6). In addition, enterovirus D68 and various rhinoviruses are linked to pneumoniae (8).

RNA viruses, including enteroviruses, are characterized by high mutation rates, leading to the generation of new genetic variants. In addition, these viruses can undergo genome recombination, a significant driving force in their evolution. Recombination events can result in the emergence of novel enterovirus variants with increased pathogenicity and fitness, contributing to the dynamic nature of these viruses (9, 10). In particular, intertypic recombination plays a crucial role in the emergence of highly pathogenic circulating vaccine-derived polioviruses. These recombinant viruses have been responsible for numerous outbreaks of paralytic poliomyelitis worldwide (11). Enteroviruses have been found to exhibit species-specific differences (12), with *Enterovirus B* being the most detected species globally, *Enterovirus A* being a more common species in Asia, and *Enterovirus C* being more prevalent in Africa. Although outbreaks typically display serotype-specific dominance, it is important to note that the co-circulation of multiple serotypes can occur worldwide (13–16). To gain a comprehensive understanding of enteroviruses, it is crucial to obtain complete enterovirus genome sequences. These genome sequences provide valuable insights into the diversity, epidemiology, evolution, and pathogenicity of enteroviruses. However, achieving complete genome sequencing can be challenging due to their high variability and dynamic nature in epidemics. Challenges arise in primer design to ensure coverage of diverse strains and in managing the sequencing costs associated with large-scale genomic studies (2, 17, 18).

In recent years, the MinION sequencer from Oxford Nanopore Technologies (ONT) has gained widespread popularity for viral genome sequencing (18–24). This is primarily attributed to its portability and cost-effectiveness compared to other sequencing technologies. In our previous study (2), we employed Illumina sequencing technology to sequence the genome of 52 enterovirus isolates. However, the cost associated with constructing genome libraries using this method was not sufficiently low to make it suitable for widespread use in virus surveillance. In this study, we aimed to address this limitation by providing a cost-effective protocol for generating complete enterovirus genomes using a MinION sequencer. Our protocol also enables the generation of polyprotein sequences and peptide sequences of the uORF. Through comprehensive phylogenetic analyses of the whole genome and uORF, we investigated the epidemics of enterovirus A71 subgenogroups C1 and B5 circulating in Taiwan in 2019 and in Malaysia in 2018.