AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Genome Sequences of 16 Enterovirus Isolates from Environmental Sewage in Guatemala, 2019 to 2021

Chelsea Harrington,[a] Leanna Sayyad,[b] Christina Castro,[b] Jamaica Hill,[c] Stacey Jeffries-Miles,[a] Hanen Belgasmi,[a] Gloria Rey-Benito,[d] María Linda Mendoza Prillwitz,[e] Leticia Castillo Signor,[f] Nancy Gerloff[a]

[a]Division of Viral Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA
[b]Cherokee Nation Assurance, Atlanta, Georgia, USA
[c]IHRC Inc., Atlanta, Georgia, USA
[d]Pan American Health Organization, World Health Organization, Washington, DC, USA
[e]National Health Laboratory Guatemala, Villa Nueva, Guatemala
[f]Ministry of Public Health and Social Assistance, Guatemala City, Guatemala

ABSTRACT Enteroviruses can cause human infectious disease. We report 16 near-complete genome sequences of enteroviruses that were isolated through environmental surveillance of wastewater in Guatemala.

The genus *Enterovirus* contains 15 species and belongs to the *Picornaviridae* family, a large family of nonenveloped positive-sense, single-stranded RNA viruses. The *Enterovirus* B (EV-B) species contains 63 serotypes and is the largest EV species (1). The EV-C species contains 23 serotypes, which includes the three polioviruses (2). Ten EV-B (1 coxsackievirus [CV] type B5, 2 echovirus type 1 [E-1], 1 E-3, 1 E-7, 2 E-11, 1 E-25, 1 E-29, and 1 E-33) and six EV-C (3 CV A13, 1 CV A20, 1 CV A24, and 1 EV C99) were identified through isolation and genome sequencing from environmental sewage collected in Villa Nueva (VNA; GPS coordinates 14.5269 to 90.5875) and San Juan Sacatepéquez, Guatemala (SJS; GPS 14.7236, 90.6520) from 2019 to 2021 (Table 1).

Sewage samples were processed using the concentration and filter elution (CaFÉ) method, as described previously (3, 4). Resulting concentrates were inoculated into cells for enterovirus isolation according to the World Health Organization protocol (5). Briefly, concentrates were inoculated into rhabdomyosarcoma (RD) cells and incubated for 5 days at 37°C. On day 5, the cells were observed for cytopathic effect (CPE).

Viral RNA was extracted from CPE-positive cell culture supernatants using the MagMAX pathogen RNA/DNA kit on a KingFisher Flex system (Thermo Fisher Scientific). Viral RNA was amplified using a sequence-independent, single-primer amplification (SISPA) protocol (6–8). Viral RNA was reverse transcribed using SuperScript III reverse transcriptase (Thermo Fisher Scientific) and a 28-base primer with eight random nucleotides on the 3′ end (CCTTGAAGGC GGACTGTGAGNNNNNNNN). A complementary strand was synthesized using the Klenow fragment of DNA polymerase I (New England BioLabs). Illumina libraries were prepared using the Nextera XT library preparation kit on 69 pooled samples. The samples were sequenced on an Illumina MiSeq system using a 500-cycle paired-end run as previously described (9).

A custom in-house bioinformatics pipeline (10) was used to process raw FASTQ data and for *de novo* assembly of each isolate's read. Within the pipeline, multiple preprocessing steps were conducted before the FASTQ reads were assembled. First, the host data were removed using default parameters in Bowtie 2 v2.3.3.1 (11–13), followed by primer trimming, adapter trimming, and Phred quality score filtering using Cutadapt v2.3 (parameters for filtering: reads with a quality score of $<$20, read length of $<$50 nucleotides, and error rates of $>$0.15) (14), and finally duplicate reads were removed using the Dedup.py script in Python (15). Deduplicated reads were *de novo* assembled into contigs using default parameters in SPAdes v3.15.0 (16).

Editor Jelle Matthijnssens, KU Leuven

Address correspondence to Nancy Gerloff, NGerloff@cdc.gov.

The authors declare no conflict of interest.

Received 13 June 2022
Accepted 25 July 2022
Published 11 August 2022

TABLE 1 Sequencing summary and characteristics of 16 enteroviruses from Guatemala, 2019 to 2021

Isolate	Virus	Taxonomy	Collection date (mm/dd/yyyy)	Collection site	GenBank accession no.	Total no. of reads[a]	Length (bp)	GC content (%)
A549-010	Coxsackievirus B5	*Enterovirus B*	11/22/2019	VNA	OL955504	9,553	7,302	47.7
HLF-000	Echovirus 3	*Enterovirus B*	11/20/2019	SJS	OL955506	15,035	7,342	47.4
HLF-006	Coxsackievirus A13	*Enterovirus C*	11/22/2019	SJS	OL955507	23,120	7,395	44.8
MA104-000	Echovirus 1	*Enterovirus B*	11/20/2019	SJS	OL955509	6,015	7,132	47.0
MA104-002	Echovirus 7	*Enterovirus B*	11/20/2019	SJS	OL955511	14,881	7,270	47.6
RD-000	Echovirus 29	*Enterovirus B*	11/20/2019	SJS	OL955512	4,717	7,314	47.8
169-41CQU3372	Echovirus 25	*Enterovirus B*	09/16/2020	SJS	ON383153	9,325	7,259	47.6
179-51CBM4841	Echovirus 11	*Enterovirus B*	07/13/2020	SJS	ON383154	35,931	7,312	47.5
183-55CBM2468	Coxsackievirus A13	*Enterovirus C*	06/10/2020	SJS	ON383155	17,008	7,355	44.4
183-55CBM2468-1	Coxsackievirus A24	*Enterovirus C*	06/10/2020	SJS	ON383156	13,212	7,365	44.7
190-62ACB0312	Enterovirus C99	*Enterovirus C*	05/15/2020	SJS	ON383157	8,644	7,302	44.9
129-1CBM1352	Echovirus 33	*Enterovirus B*	09/01/2021	SJS	ON383146	33,169	7,240	47.9
145-17ACB0328	Coxsackievirus A20	*Enterovirus C*	05/18/2021	SJS	ON383147	4,154	7,185	45.8
146-18PLA0330	Coxsackievirus A13	*Enterovirus C*	05/18/2021	VNA	ON383149	57,563	7,318	45.1
148-20CQU0199	Echovirus 11	*Enterovirus B*	04/16/2021	SJS	ON383150	7,861	7,182	47.7
157-29CVP8542	Echovirus 1	*Enterovirus B*	01/25/2021	VNA	ON383152	3,399	7,211	47.3

[a] Number of reads after quality control and deduplication.

Consensus genome sequences were verified through read mapping, BLAST alignments using MAFFT, and annotations using Geneious vR11.

The 16 near-complete genome sequences ranged from 7,132 to 7,395 bp in length. Their GC content was between 44.4% and 47.9%, and the median read coverage was 11,382 (interquartile range [IQR], 7,399 to 18,536). These genome sequences share 80 to 90% pairwise identity to previously submitted nucleotide sequences and, therefore, are distinct from other enterovirus genomes in GenBank.

Data availability. The 16 EVs have been submitted to GenBank, and the raw sequencing reads have been deposited in the Sequence Read Archive under BioProject PRJNA835862. All accession numbers are reported in Table 1.

ACKNOWLEDGMENTS

We thank Rachel Marine and Anna Montmayeur for their technical assistance.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official positions of the Centers for Disease Control and Prevention.

REFERENCES

1. Lugo D, Krogstad P. 2016. Enteroviruses in the early 21st century: new manifestations and challenges. Curr Opin Pediatr 28:107–113. https://doi.org/10.1097/MOP.0000000000000303.
2. Brown B, Oberste MS, Maher K, Pallansch MA. 2003. Complete genomic sequencing shows that polioviruses and members of human enterovirus species C are closely related in the noncapsid coding region. J Virol 77:8973–8984. https://doi.org/10.1128/JVI.77.16.8973-8984.2003.
3. Alleman MM, Coulliette-Salmond AD, Wilnique P, Belgasmi-Wright H, Sayyad L, Wong K, Gue E, Barrais R, Rey-Benito G, Burns CC, Vega E. 2021. Environmental surveillance for polioviruses in Haïti (2017–2019): the dynamic process for the establishment and monitoring of sampling sites. Viruses 13:505. https://doi.org/10.3390/v13030505.
4. Belgasmi H, Miles SJ, Sayyad L, Wong K, Harrington C, Gerloff N, Coulliette-Salmond AD, Guntapong R, Tacharoenmuang R, Ayutthaya AIN, Apostol LNG, Valencia MA-LD, Burns CC, Benito G-R, Vega E. 2022. CaFÉ: a sensitive, low-cost filtration method for detecting polioviruses and other enteroviruses in residual waters. Front Environ Sci 10. https://doi.org/10.3389/fenvs.2022.914387.
5. World Health Organization Global Polio Eradication Initiative. 2015. Guidelines on environmental surveillance for detection of polioviruses. https://polioeradication.org/wp-content/uploads/2016/07/GPLN_GuidelinesES_April2015.pdf. Accessed 29 April 2022.
6. Reyes GR, Kim JP. 1991. Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. Mol Cell Probes 5:473–481. https://doi.org/10.1016/S0890-8508(05)80020-9.
7. Ng TFF, Kondov NO, Deng X, Van Eenennaam A, Neibergs HL, Delwart E. 2015. A metagenomics and case-control study to identify viruses associated with bovine respiratory disease. J Virol 89:5340–5349. https://doi.org/10.1128/JVI.00064-15.
8. Ng TFF, Marine R, Wang C, Simmonds P, Kapusinszky B, Bodhidatta L, Oderinde BS, Wommack KE, Delwart E. 2012. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. J Virol 86:12161–12175. https://doi.org/10.1128/JVI.00869-12.
9. Montmayeur AM, Ng TFF, Schmidt A, Zhao K, Magaña L, Iber J, Castro CJ, Chen Q, Henderson E, Ramos E, Shaw J, Tatusov RL, Dybdahl-Sissoko N, Endegue-Zanga MC, Adeniji JA, Oberste MS, Burns CC. 2017. High throughput next-generation sequencing of polioviruses. J Clin Microbiol 55:606–615. https://doi.org/10.1128/JCM.02121-16.
10. Wagner DD, Marine RL, Ramos E, Ng T, Castro CJ, Okomo-Adhiambo M, Harvey K, Doho G, Kelly R, Jain Y, Tatusov RL, Silva H, Rota PA, Khan AN, Oberste MS. 2022. VPipe: an automated bioinformatics platform for assembly and management of viral next-generation sequencing data. Microbiol Spectr 10:e0256421. https://doi.org/10.1128/spectrum.02564-21.
11. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. https://doi.org/10.1186/gb-2009-10-3-r25.
12. Langmead B, Wilks C, Antonescu V, Charles R. 2018. Scaling read aligners to hundreds of threads on general-purpose processors. Bioinformatics 35:421–435. https://doi.org/10.1093/bioinformatics/bty648.

13. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
14. Martin M. 2011. Cutadapt removes adapter sequences from high throughput sequencing reads. EMBnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200.
15. Deng X, Naccache SN, Ng T, Federman S, Li L, Chiu CY, Delwart EL. 2015. An ensemble strategy that significantly improves de novo assembly of microbial genomes from metagenomic next-generation sequencing data. Nucleic Acids Res 43:e46. https://doi.org/10.1093/nar/gkv002.
16. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.