AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

| Virology | Announcement

Coding-complete genome of human cytomegalovirus extracted from a nasopharyngeal swab in Brazil

Amanda Mendes Silva Cruz,[1,2] Edivaldo Costa Sousa, Jr,[3] Luana Soares Barbagelata,[1] Wanderley Dias das Chagas, Jr,[1] Agatha Monike Silva Nunes,[1] Miriam Teresinha Furlam Prando Livorati,[4] Walquiria Aparecida Ferreira de Almeida,[5] João Farias da Trindade,[6] Andreia Santos Costa,[7] Fernando Neto Tavares,[1] Maísa Silva de Sousa,[2] Mirleide Cordeiro dos Santos[1]

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT We report the coding-complete genome sequence of *Cytomegalovirus humanbeta5* (BR/2023-012844/IEC/NS/2023), recovered from a nasopharyngeal swab collected from a 5-month-old infant with severe acute respiratory syndrome during a respiratory syncytial virus outbreak in Brazil.

KEYWORDS *Cytomegalovirus humanbeta5*, Brazil

Human cytomegalovirus (HCMV) is a member of the family *Herpesviridae*, subfamily *Betaherpesvirinae*, and genus *Cytomegalovirus* (1). Its genome consists of a double-stranded DNA molecule of >200 kbp (2). HCMV genomes are most commonly recovered from blood, urine, and saliva, particularly in congenital infections and in immunocompromised individuals (3–7). Recovering viral genomes from upper respiratory tract samples, such as nasopharyngeal swabs, is uncommon because this anatomical site is not routinely examined for HCMV in clinical or surveillance settings. However, obtaining the viral genome from the nasopharynx may provide key insights into viral shedding, reactivation, and coinfection dynamics. Shotgun metagenomics is well suited for recovering viral genomes directly from clinical samples (3). Notably, as of 1 September 2025, only partial HCMV genome sequences from Brazil were available in GenBank (8).

The nasopharyngeal swab analyzed in this study (2023-012844/IEC) was obtained from a 5-month-old infant hospitalized in an intensive care unit in Macapá, Amapá, located in the Brazilian Amazon. The GPS coordinates for sample collection were 0.0388900 and −51.0663900. The sample was collected in April 2023 by the Health Surveillance Secretariat during a respiratory syncytial virus (RSV) outbreak in Brazil. Viral DNA was extracted using the Extracta Kit – Pathogen DNA and RNA MDx, which employs magnetic microparticle technology, following the manufacturer's instructions. Library preparation and sequencing were performed at the Respiratory Virus Laboratory of the Evandro Chagas Institute. Libraries were constructed using the Illumina Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) and sequenced on the Illumina NextSeq 550 System with 2 × 150 bp paired-end reads, generating 39,712,796 read pairs. Quality filtering of reads in FASTQ format was performed using fastp v.0.23.1 (9), and *de novo* assembly was conducted using MEGAHIT v.1.2.9 (10) (Table 1). Taxonomic classification of contigs was carried out using Kraken2 v.2.1.0 (11) with the Standard-16 database, which identified multiple contigs assigned to *Cytomegalovirus*. Classification outputs were visualized using the Pavian R package v.1.2 (12). Manual review and open reading frame prediction were performed in Geneious Prime (13), where multiple overlapping contigs were identified. These contigs, supported by read mapping and alignment with MAFFT (14), allowed reconstruction of a coding-complete genome

Editor Jelle Matthijnssens, Katholieke Universiteit Leuven, Leuven, Belgium

Address correspondence to Fernando Neto Tavares, fernandotavares@iec.gov.br.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 1 September 2025
Accepted 11 December 2025
Published 13 January 2026

TABLE 1 Sequencing and genome metrics

Metric	Data for BR/2023-012844/IEC/NS/2023
No. of raw reads	39,712,796
No. of reads after trimming	36,229,206
Average read length (bp)	158
Total no. of contigs	912,240
No. of contigs related to *Cytomegalovirus humanbeta5*	16
Genome size (bp)	231,043
GC content (%)	57.5
Coverage (×)	801.4
N50	466
GenBank accession no.	PX246278

relative to the reference strain (GenBank accession number PP803491). Each contig was additionally confirmed by BLASTn (15).

The 2023-012844_CMV/IEC genome exhibited a typical HCMV genomic organization, with a length of 231,043 bp and 178 predicted genes. The mean read coverage was 801.4×, and the GC content (57.5%) was consistent with other *Cytomegalovirus humanbeta5* genomes available in GenBank. The genome displayed high nucleotide identity (>98%) with related HCMV strains available in GenBank (KJ361950; KJ361951; KJ426589; KT726951; KY490088; MT044485), demonstrating the genetic similarity of the 2023-012844/IEC strain to previously characterized isolates. The shotgun metagenomic approach provided evidence of HCMV genomic material in the nasopharynx of an RSV-positive infant with severe acute respiratory syndrome. As RSV infection can modulate host immune responses (16), further investigations are warranted to elucidate viral coinfections, HCMV reactivation, and the potential impact of congenital HCMV infection.

ACKNOWLEDGMENTS

We thank the technical professional from the Health Department of the State of Amapá for collecting the sample, the Central Laboratory of the State of Amapá for sending the sample, and the Evandro Chagas Institute, Ministry of Health, for the financial resources and physical infrastructure to carry out this work.

AUTHOR AFFILIATIONS

[1]Laboratory of Respiratory Viruses, Section of Virology, Evandro Chagas Institute, Ministry of Health, Ananindeua, Pará, Brazil
[2]Graduate Program in Pathology of Tropical Diseases, Federal University of Pará, Belém, Brazil
[3]Labepileish, Section of Parasitology, Evandro Chagas Institute, Ministry of Health, Ananindeua, Pará, Brazil
[4]General Coordination of Public Health Laboratories, Brazilian Ministry of Health, Brasília, Brazil
[5]Secretariat of Health and Environmental Surveillance, Brazilian Ministry of Health, Brasília, Brazil
[6]Secretariat for Health of Amapá State, Macapá, Brazil
[7]Central Public Health Laboratory of the Amapá State, Brazilian Ministry of Health, Macapá, Brazil

AUTHOR ORCIDs

Amanda Mendes Silva Cruz http://orcid.org/0000-0002-9513-9317
Edivaldo Costa Sousa, Jr http://orcid.org/0000-0002-5423-238X
Luana Soares Barbagelata http://orcid.org/0000-0003-3586-0601

Wanderley Dias das Chagas, Jr http://orcid.org/0000-0002-9417-0742
Agatha Monike Silva Nunes http://orcid.org/0000-0001-8267-5498
Miriam Teresinha Furlam Prando Livorati http://orcid.org/0009-0005-4454-8398
Walquiria Aparecida Ferreira de Almeida http://orcid.org/0000-0003-2082-6283
João Farias da Trindade http://orcid.org/0009-0003-7558-0900
Andreia Santos Costa http://orcid.org/0009-0004-6557-7262
Fernando Neto Tavares http://orcid.org/0000-0002-1385-5169
Maísa Silva de Sousa http://orcid.org/0000-0001-9944-7838
Mirleide Cordeiro dos Santos http://orcid.org/0000-0001-8140-4688

FUNDING

Funder	Grant(s)	Author(s)
Foundation for Scientific and Technological Development in Health	DIREX-003-FIO-24	Luana Soares Barbagelata Wanderley Dias das Chagas Jr Agatha Monike Silva Nunes
Conselho Nacional de Desenvolvimento Científico e Tecnológico	159996/2021-0	Amanda Mendes Silva Cruz

AUTHOR CONTRIBUTIONS

Amanda Mendes Silva Cruz, Data curation, Methodology, Software, Writing – original draft | Edivaldo Costa Sousa, Jr, Data curation, Methodology, Software, Validation | Luana Soares Barbagelata, Methodology | Walquiria Aparecida Ferreira de Almeida, Investigation | Fernando Neto Tavares, Funding acquisition, Supervision, Writing – review and editing.

DATA AVAILABILITY

The whole genome sequencing project of sample BR/2023-012844/IEC/NS/2023 was deposited in GenBank under accession number PX246278. The accession number for the raw shotgun metagenomic sequencing reads in the NCBI Sequence Read Archive is SRR35176174. This data set is associated with BioProject accession number PRJNA1312120 and BioSample accession number SAMN50854041 (2023-012844/IEC).

ETHICS APPROVAL

The study was approved by the Research Ethics Committee of the Evandro Chagas Institute, Pará, Brazil (approval number 79631424.6.0000.0019).

REFERENCES

1. International Committee on Taxonomy of Viruses. 2011. In ICTV ninth report: 2009 taxonomy release. https://ictv.global/report_9th/dsDNA/Herpesvirales/Herpesviridae.
2. Murphy E, Shenk TE. 2008. Human Cytomegalovirus, p 1–19. In Human cytomegalovirus. Berlin, Heidelberg.
3. Tsai HP, Yeh CS, Lin IT, Ko WC, Wang JR. 2020. Increasing cytomegalovirus detection rate from respiratory tract specimens by a new laboratory-developed automated molecular diagnostic test. Microorganisms 8:1063. https://doi.org/10.3390/microorganisms8071063
4. Griffiths P, Reeves M. 2021. Pathogenesis of human cytomegalovirus in the immunocompromised host. Nat Rev Microbiol 19:759–773. https://doi.org/10.1038/s41579-021-00582-z
5. Barroso Hernández S, Rodríguez Sabillón JA, Álvarez López Á, García de Vinuesa Calvo E, Calvo Cano A, Robles Pérez-Monteoliva NR. 2023. Enfermedad por citomegalovirus en los pacientes inmunodeprimidos por glomerulonefritis necrosante pauciinmune. Medicina Clínica 160:199–202. https://doi.org/10.1016/j.medcli.2022.07.009
6. Giorgione V, Krajden Haratz K, Gull I, Brusilov M, Birnbaum R, Blecher Y, Malinger G, Kaplan A, Beer G, Kapusta L. 2023. Myocardial function in fetuses with congenital cytomegalovirus infection. Fetal Diagn Ther 50:430–437. https://doi.org/10.1159/000533280
7. Ozdemir E, Sarac Sivrikoz T, Sarsar K, Tureli D, Onel M, Demirci M, Yapar G, Yurtseven E, Has R, Agacfidan A, Kirkoyun Uysal H. 2024. Evaluation of congenital cytomegalovirus infection in pregnant women admitted to a university hospital in Istanbul. Viruses 16:414. https://doi.org/10.3390/v16030414
8. National Center for Biotechnology Information. 2025. NCBI virus sequences for discorevy. Available from: https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Nucleotide&VirusLineage_ss=Human%20betaherpesvirus%205,%20taxid:10359&utm_source=nuccore&utm_medium=referral
9. Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560
10. Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, Yamashita H, Lam T-W. 2016. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods 102:3–11. https://doi.org/10.1016/j.ymeth.2016.02.020

11. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. Genome Biol 20:257. https://doi.org/10.1186/s13059-019-1891-0
12. Breitwieser FP, Salzberg SL. 2020. Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. Bioinformatics 36:1303–1304. https://doi.org/10.1093/bioinformatics/btz715
13. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199
14. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010
15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
16. Heinonen S, Velazquez VM, Ye F, Mertz S, Acero-Bedoya S, Smith B, Bunsow E, Garcia-Mauriño C, Oliva S, Cohen DM, Moore-Clingenpeel M, Peeples ME, Ramilo O, Mejias A. 2020. Immune profiles provide insights into respiratory syncytial virus disease severity in young children. Sci Transl Med 12:eaaw0268. https://doi.org/10.1126/scitranslmed.aaw0268