Whole-Genome Sequencing of an Unusual Human Papillomavirus (HPV71) from Latin America (Brazil)

Edivaldo Costa Sousa Junior,ª Allan Kaio Silva,ª Laryssa Danielle da Silva Reis,ª Lana Patricia da Silva Fonseca,ª Fabiano Reis da Silva,ᵇ Fabiolla da Silva dos Santos,ª Jessylene Almeida Ferreira,ª Clayton Pereira Silva de Lima,ᵇ Michelle Carvalho de Abreu,ª Rodrigo Vellasco Duarte Silvestre,ª Marcio Roberto Teixeira Nunesᵇ

ªSection of Virology, Evandro Chagas Institute, Ministry of Health, Ananindeua, Brazil
ᵇCenter of Technological Innovation, Evandro Chagas Institute, Ministry of Health, Ananindeua, Brazil

ABSTRACT We report the complete genome sequencing of human papillomavirus 71 from Latin America (Brazil).

The *Papillomaviridae* family is composed of 53 genera, and 5 of them (*Alphapapillomavirus*, *Betapapillomavirus*, *Gammapapillomavirus*, *Mupapillomavirus*, and *Nupapillomavirus*) can infect humans (1). Within the *Alphapapillomavirus* genus, human papillomavirus 71 (HPV71) is associated with mucosal basal keratinocyte infections (2). The genome is a double-stranded circular DNA with a size of 8,017 to 8,038 kb and eight genes enclosed in a nonenveloped icosahedral capsid structure with a diameter of 52 to 55 nm (3). In Brazil, the epidemiology of HPV71 is still unclear, with inconclusive descriptions in the literature even about its relationship to cancer (4–6).

Here, we describe the complete genome sequencing of an infection identified by molecular tools in a Brazilian patient (59 years of age, from the Amazon region) with inflammatory cytology. The genomic DNA was extracted from a clinical sample (cervical smear) that had been collected from a 59-year-old female patient and stored in a solution buffer until DNA extraction using a QIAamp DNA extraction kit (Qiagen, Germany). The genomic library was prepared using the Nextera XT DNA sample preparation kit (Illumina, USA). The quality of the library was verified using a Bioanalyzer 2010 (Agilent Technologies), and the library was sequenced on a HiSeq 2500 instrument (Illumina, USA) with a 2 × 100-bp paired-end format sequencing kit v.4. All laboratory procedures were performed according to the manufacturer's instructions. The DNA sequencing generated 40,278,934 reads, which were assembled using a *de novo* strategy with MEGAHIT software v.1.2.9 (7), producing 140,912 contigs. These contigs were taxonomically annotated with Kraken v.1 (8), and 10 were related to *Alphapapillomavirus 14* (HPV71), presenting 98.4% nucleotide identity to the HPV71 reference (GenBank accession number NC_039089). The contigs were mapped to the reference sequence (NC_039089) to generate a scaffold. The scaffold generated was about 8,041 kb and was used for reference mapping. We performed reference mapping with SOAP3-dp (9) and generated a consensus sequence using BCFtools (10). The prediction of open reading frames (ORFs) and functional annotation were automatically performed with Geneious v.8.1.9 (11) (similarity, 94%) using a sequence database with 660 annotated papillomavirus genomes retrieved from the Papillomavirus Episteme (PaVE) (12). The genome was manually curated by comparison of the coding ORFs with these genomes using Geneious v.8.1.9 (11). Our final result was the assembly of a genome that is similar to that of the closest species, isolate Qv21030 (AY330620), with 99.84% nucleotide identity, determined using Geneious v.8.1.9 (11). All tools were used with default parameters unless otherwise specified. The genome showed a circular double-stranded structure about 8,033 bp long, with eight fully identified genes and a

GC content of 44.4%. The genome identified had a mean coverage of about 25.1× (ranging from 6× to 191×), with the reads overlapping at both ends, forming a circular contig containing the complete genome sequence.

The genome obtained will allow us to contribute to genomic studies with other HPV71 isolates already described throughout the world and will facilitate better understanding of the pathogenic and epidemiological aspects of HPVs in Latin America.

Data availability. The complete genome sequence for Brazilian HPV71 has been deposited in GenBank under the accession number MT250602 and in the SRA under accession number SRR11440049.

REFERENCES

1. Van Doorslaer K, Chen Z, Bernard H, Chan PKS, DeSalle R, Dillner J, Forslund O, Haga T, McBride AA, Villa LL, Burk RD, ICTV Report Consortium. 2018. ICTV virus taxonomy profile: Papillomaviridae. J Gen Virol 99:989–990. https://www.microbiologyresearch.org/content/journal/jgv/10.1099/jgv.0.001105.
2. Matsukura T, Sugase M. 2001. Relationships between 80 human papillomavirus genotypes and different grades of cervical intraepithelial neoplasia: association and causality. Virology 283:139–147. https://doi.org/10.1006/viro.2001.0865.
3. Modis Y, Benes LT, Stephen CH. 2002. Atomic model of the papillomavirus capsid. EMBO J 21:4754–4762. https://doi.org/10.1093/emboj/cdf494.
4. Oliveira GR, Vieira VC, Barral MFM, Dowich V, Soares MA, Concalves CV, Martinez A. 2013. Fatores de risco e prevalência da infecção pelo HPV em pacientes de Unidades Básicas de Saúde e de um Hospital Universitário do Sul do Brasil. Rev Bras Ginecol Obstet 35:226–232. https://doi.org/10.1590/S0100-72032013000500007.
5. Brown CR, Leon ML, Munoz K, Fagioni A, Amador LG, Frain B, Tu W, Qadadri B, Brown DR. 2009. Human papillomavirus infection and its association with cervical dysplasia in Ecuadorian women attending a private cancer screening clinic. Braz J Med Biol Res 42:629–636. https://doi.org/10.1590/s0100-879x2009000700007.
6. Bruni L, Albero G, Serrano B, Mena M, Gómez D, Muñoz J, Bosch FX, de Sanjosé S, ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre). Human papillomavirus and related diseases in the world. Summary Report 22 January 2019. https://hpvcentre.net/statistics/reports/BRA.pdf.
7. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033.
8. Wood D-E, Salzberg S-L. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:R46. https://doi.org/10.1186/gb-2014-15-3-r46.
9. Luo R, Wong T, Zhu J, Liu C-M, Zhu X, Wu E, Lee L-K, Lin H, Zhu W, Cheung DW, Ting H-F, Yiu S-M, Peng S, Yu C, Li Y, Li R, Lam T-K. 2013. SOAP3-dp: fast, accurate and sensitive GPU-based short read aligner. PLoS One 8:e65632. https://doi.org/10.1371/journal.pone.0065632.
10. Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. 2016. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics 32:1749–1751. https://doi.org/10.1093/bioinformatics/btw044.
11. 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.
12. Van Doorslaer K, Li Z, Xirasagar S, Maes P, Kaminsky D, Liou D, Sun Q, Kaur R, Huyen Y, McBride AA. 2017. The Papillomavirus Episteme: a major update to the papillomavirus sequence database. Nucleic Acids Res 45:D499–D506. https://doi.org/10.1093/nar/gkw879.