




Small Circular Rep-Encoding Single-Stranded DNA Genomes in Peruvian Diarrhea Virome

Eda Altan,^{a,b} Juana Del Valle Mendoza,^{c,d} Xutao Deng,^{a,b} Tung G. Phan,^{a,b} Mohammadreza Sadeghi,^{a,b,e}  Eric L. Delwart^{a,b}

Blood Systems Research Institute, San Francisco, California, USA^a; Department of Laboratory Medicine, University of California at San Francisco, San Francisco, California, USA^b; School of Medicine, Research and Innovation Centre of the Faculty of Health Sciences, Universidad Peruana de Ciencias Aplicadas, Lima, Peru^c; Instituto de Investigación Nutricional, Lima, Peru^d; Department of Virology, University of Helsinki, Helsinki, Finland^e

ABSTRACT Metagenomic analysis of diarrhea samples revealed the presence of numerous human enteric viruses and small circular Rep-encoding single-stranded DNA (CRESS-DNA) genomes. One such genome was related to smacoviruses, while eight others were related to genomes reported in the feces of different mammals. The tropism of these CRESS-DNA viruses remains unknown.

Fecal samples from 300 Peruvian patients with diarrhea were pooled in groups of 10 and processed to enrich for viral particle-associated nucleic acids using filtration and nuclease treatment. Amplification of nucleic acids using random reverse transcription (RT)-PCR was followed by use of the Illumina Nextera method to generate sequencing libraries, which were sequenced using the MiSeq instrument (250-bp paired-end reads) (1). *De novo* assembly was performed using the Ensemble program (2). The resulting singlets and contigs were analyzed using BLASTx to search for similarity to viral proteins in GenBank's Virus RefSeq.

Numerous human viruses were identified with E scores of $<10^{-30}$, including anelloviruses, adenovirus D, *Caliciviridae* (Sapporo virus, Norwalk virus), *Picornaviridae* (enterovirus A/B/C, parechovirus, aichivirus A, cosavirus C/D/E, salivirus, cardiovirus), *Astroviridae* (mamastrovirus 1, astrovirus MLB1), *Parvoviridae* (bocavirus 1, bufavirus 3, adeno-associated virus), and picobirnavirus. Other viruses of unknown origin were also identified, including husavirus and human plasma-associated gemycircular virus (3, 4). Viral sequences of dietary origins were identified, including members of the plant-infecting *Virgaviridae* and *Betaflexiviridae* families and chicken anemia virus.

Also detected were reads with translated protein similarity to the Rep protein of circular Rep-encoding single-stranded DNA (CRESS-DNA) viral genomes (5), a large and diverse viral group that infects diverse eukaryote hosts, including vertebrates, fungi, plants, and insects (6–8), and is often found in environmental samples (8–10). A member of that clade (SsHADV) was shown to infect both a fungus and insects (11).

Here, the complete genome sequences of nine CRESS-DNA genomes were assembled using the contigs generated by the Ensemble program and the Geneious program version R10.1.2 (Auckland, New Zealand) to identify overlapping reads until direct repeats were generated at both extremities, which were then circularized into CRESS-DNA genomes.

Genomes were 2,548 to 2,853 bases long and encoded Rep and Cap proteins in opposite orientations. One genome was closely related to those of CRESS-DNA smacoviruses previously reported in fecal samples from diverse mammals (12, 13). The closest relatives were from captive chimpanzee (GenBank accession no. KT600069) and

Received 6 July 2017 Accepted 8 August 2017 Published 21 September 2017

Citation Altan E, Del Valle Mendoza J, Deng X, Phan TG, Sadeghi M, Delwart EL. 2017. Small circular Rep-encoding single-stranded DNA genomes in Peruvian diarrhea virome. *Genome Announc* 5:e00822-17. <https://doi.org/10.1128/genomeA.00822-17>.

Copyright © 2017 Altan et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Eric L. Delwart, delwarte@medicine.ucsf.edu.

human (NC_026318) feces, with 95.4 to 96.2% Rep and 68.1 to 79.5% Cap identities. A similar tentative stem loop origin of replication was located 3' of the *rep* gene.

The remaining genomes were closely related to each other, with Rep and Cap identities of 71.5 to 97.5% and 33.8 to 98.6%, respectively. The closest relatives in GenBank were CRESS-DNA virus from rhesus macaque feces (KU043400), dromedary stool-associated circular single-stranded DNA (ssDNA) virus (KM573765) (14), and bovine feces-associated circular DNA virus 1 (NC_030130) (15), with Rep protein identities ranging from 72.9 to 93.7%, 47.5 to 54.8%, and 44.8 to 49.5% and Cap protein identities ranging from 33.6 to 42.1%, 20.8 to 24.4%, and 22.5 to 27.4%, respectively. The intergenic regions 3' of the *rep* gene all contained a stem loop with a potential origin consensus nanomer of NANNTTAC. Rolling circle replication motifs I, II, and III and superfamily 3 helicase Walker A, B, and C motifs (7) were identified in all nine genomes. Based on gene orientations and location of the stem loop, all genomes correspond to type IV CRESS-DNA viruses (7).

The cellular host(s) or any pathogenic role for these feces-associated CRESS-DNA viruses remains unknown. Potential cell hosts include human enteric cells and ingested plant and/or animal food products.

Accession number(s). The complete genome sequences of hudisaviruses and human feces smacovirus 3 isolate P21 have been deposited at GenBank under the accession no. [MF351513](https://doi.org/10.1093/genbank/MF351513) through [MF351520](https://doi.org/10.1093/genbank/MF351520) and [MF347417](https://doi.org/10.1093/genbank/MF347417), respectively.

ACKNOWLEDGMENT

We thank the Blood Systems Research Institute for support.

REFERENCES

- Li L, Deng X, Mee ET, Collot-Teixeira S, Anderson R, Schepelmann S, Minor PD, Delwart E. 2015. Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. *J Virol Methods* 213:139–146. <https://doi.org/10.1016/j.jviromet.2014.12.002>.
- 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>.
- Oude Munnink BB, Cotten M, Deijns M, Jebbink MF, Bakker M, Farsani SM, Canuti M, Kellam P, van der Hoek L. 2015. A novel genus in the order *Picornavirales* detected in human stool. *J Gen Virol* 96:3440–3443. <https://doi.org/10.1099/jgv.0.000279>.
- Zhang W, Li L, Deng X, Blümel J, Nübling CM, Hunfeld A, Baylis SA, Delwart E. 2016. Viral nucleic acids in human plasma pools. *Transfusion* 56:2248–2255. <https://doi.org/10.1111/trf.13692>.
- Rosario K, Dayaram A, Marinov M, Ware J, Kraberger S, Stainton D, Breitbart M, Varsani A. 2012. Diverse circular ssDNA viruses discovered in dragonflies (*Odonata: Epirocta*). *J Gen Virol* 93:2668–2681. <https://doi.org/10.1099/vir.0.045948-0>.
- Delwart E, Li L. 2012. Rapidly expanding genetic diversity and host range of the *Circoviridae* viral family and other Rep encoding small circular ssDNA genomes. *Virus Res* 164:114–121. <https://doi.org/10.1016/j.virusres.2011.11.021>.
- Rosario K, Duffy S, Breitbart M. 2012. A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. *Arch Virol* 157:1851–1871. <https://doi.org/10.1007/s00705-012-1391-y>.
- Krupovic M, Ghabrial SA, Jiang D, Varsani A. 2016. *Genomoviridae*: a new family of widespread single-stranded DNA viruses. *Arch Virol* 161:2633–2643. <https://doi.org/10.1007/s00705-016-2943-3>.
- Labonté JM, Suttle CA. 2013. Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISME J* 7:2169–2177. <https://doi.org/10.1038/ismej.2013.110>.
- Dayaram A, Galatowitsch ML, Argüello-Astorga GR, van Bysterveldt K, Kraberger S, Stainton D, Harding JS, Roumagnac P, Martin DP, Lefeuvre P, Varsani A. 2016. Diverse circular replication-associated protein encoding viruses circulating in invertebrates within a lake ecosystem. *Infect Genet Evol* 39:304–316. <https://doi.org/10.1016/j.meegid.2016.02.011>.
- Liu S, Xie J, Cheng J, Li B, Chen T, Fu Y, Li G, Wang M, Jin H, Wan H, Jiang D. 2016. Fungal DNA virus infects a mycophagous insect and utilizes it as a transmission vector. *Proc Natl Acad Sci U S A* <https://doi.org/10.1073/pnas.1608013113>.
- Ng TF, Zhang W, Sachsenröder J, Kondov NO, da Costa AC, Vega E, Holtz LR, Wu G, Wang D, Stine CO, Antonio M, Mulvaney US, Muench MO, Deng X, Ambert-Balay K, Pothier P, Vinjé J, Delwart E. 2015. A diverse group of small circular ssDNA viral genomes in human and nonhuman primate stools. *Virus Evol* 1:vev017. <https://doi.org/10.1093/ve/vev017>.
- Phan TG, da Costa AC, Del Valle Mendoza J, Bucardo-Rivera F, Nordgren J, O'Ryan M, Deng X, Delwart E. 2016. The fecal virome of South and Central American children with diarrhea includes small circular DNA viral genomes of unknown origin. *Arch Virol* 161:959–966. <https://doi.org/10.1007/s00705-016-2756-4>.
- Woo PC, Lau SK, Teng JL, Tsang AK, Joseph M, Wong EY, Tang Y, Sivakumar S, Bai R, Wernery R, Wernery U, Yuen KY. 2014. Metagenomic analysis of viromes of dromedary camel fecal samples reveals large number and high diversity of circoviruses and picobirnaviruses. *Virology* 471–473:117–125. <https://doi.org/10.1016/j.virol.2014.09.020>.
- Steel O, Kraberger S, Sikorski A, Young LM, Catchpole RJ, Stevens AJ, Ladley JJ, Coray DS, Stainton D, Dayaram A, Julian L, van Bysterveldt K, Varsani A. 2016. Circular replication-associated protein encoding DNA viruses identified in the faecal matter of various animals in New Zealand. *Infect Genet Evol* 43:151–164. <https://doi.org/10.1016/j.meegid.2016.05.008>.