





Genome Sequences of Anelloviruses, Genomovirus, and Papillomavirus Isolated from Nasal Pharyngeal Swabs

Courtney L. Collins,^a Simona Kraberger,^a Rafaela S. Fontenele,^a Temitope O. C. Faleye,^b Deborah Adams,^c Sangeet Adhikari,^{b,d} Helen Sandrolini,^e Sarah Finnerty,^e Rolf U. Halden,^{b,d}  Matthew Scotch,^{b,f}  Arvind Varsani^a

^aThe Biodesign Center for Fundamental and Applied Microbiomics, School of Life Sciences, Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA

^bBiodesign Center for Environmental Health Engineering, Biodesign Institute, Arizona State University, Tempe, AZ, USA

^cBiodesign Center for Personalized Diagnostics, Biodesign Institute, Arizona State University, Tempe, AZ, USA

^dSchool of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ, USA

^eArizona State University Health Services, Arizona State University, Tempe, AZ, USA

^fCollege of Health Solutions, Arizona State University, Phoenix, AZ, USA

ABSTRACT The genome sequences of three anelloviruses (genus *Alphatorquevirus*), a genomovirus (genus *Gemykolovirus*), and an unclassified papillomavirus were identified in four human nasopharyngeal swabs, and one was positive for influenza A and one for influenza B virus. The influenza B virus-positive sample had a coinfection with an anellovirus and a papillomavirus.

Nasal pharyngeal swabs were taken from four patients with influenza-like illness as part of a routine clinical testing for influenza virus on a university campus in Arizona between February and March 2020. Two of these swabs tested positive for seasonal influenza viruses, namely, one influenza A virus (IAV) and the other influenza B virus (IBV), via rapid lateral flow immunoassay (Abbott BinaxNow). To identify associated DNA viruses in these samples, 200 μ L of the buffer from the lateral flow assay was used to extract viral DNA with the high pure viral nucleic acid kit (Roche Diagnostics, USA). Circular DNA was amplified by rolling circle amplification (RCA) using an Illustra TempliPhi kit (GE Healthcare, USA). The RCA products were used to prepare libraries using a TruSeq Nano DNA kit (Illumina, USA). The 2 \times 150-bp libraries were sequenced on a NovaSeq6000 sequencer at Psomagen Inc. (USA). The reads were trimmed using Trimmomatic v0.39 (1) and *de novo* assembled with metaSPAdes 3.14.0 (2). We identified complete genome sequences (based on terminal redundancy) of anelloviruses ($n = 3$), genomovirus ($n = 1$), and a papillomavirus ($n = 1$) using the BLASTn (3) nonredundant (nr) database online on May 2021. Reads were mapped to the viral genome sequences using BMap (4) and showed depth coverage of 24 \times to 715 \times and mapped reads per genome of 365 to 17,483 (Fig. 1A). In sample S1 (IBV positive), we found a coinfection with an anellovirus and a papillomavirus (Fig. 1).

Anelloviruses are circular single-stranded DNA (ssDNA) viruses (2.0 to 3.9 kb) (5). They are diverse and highly prevalent but are not associated with disease (6). Anellovirus genomes have three large open reading frames (ORFs). The three anelloviruses identified here are ~3.7 kb (with GC content of 52% to 53%) and share >83% nucleotide identity with those previously identified in humans (Fig. 1B and C). Based on the guidelines for the classification of mammal-infecting members of the *Anelloviridae* family (7), torque teno virus AZ1_2 (MW679003), torque teno virus AZ6_5 (MW679005), and torque teno virus AZ7_4 (MW679006) identified in the nasal pharyngeal swabs all belong to the genus *Alphatorquevirus* and species *Torque teno virus 24*, *Torque teno virus 29*, and *Torque teno virus 19*, respectively (Fig. 1C).

Genomoviruses are circular ssDNA viruses (2 to 2.4 kb) that encode bidirectionally transcribed ORFs, capsid protein (CP), and a replication associated protein (Rep), with the exception of *Fusarium graminearum* gemytripvirus 1 which is a tripartite virus (8–10). Although ubiquitous in nature, only two genomoviruses have confirmed hosts, namely,

Editor Simon Roux, DOE Joint Genome Institute

Copyright © 2022 Collins 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 Arvind Varsani, arvind.varsani@asu.edu.

The authors declare a conflict of interest.

Received 5 July 2022

Accepted 4 August 2022

Published 16 August 2022

A

Summary of viruses identified in this study

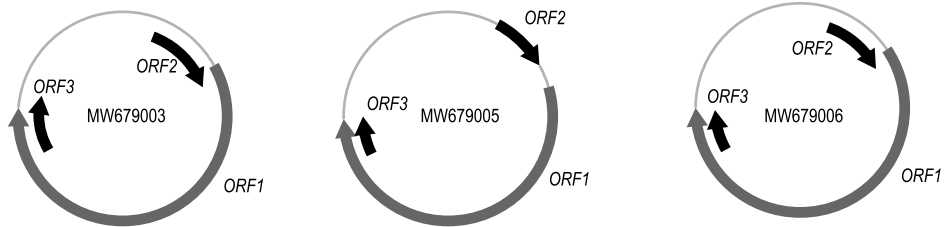
Family	Virus	Accession #	Length	% GC	S1 6-Feb-2020 S4 4-Feb-2020 S6 13-Mar-2020 S7 16-Mar-2020	Read mapping	
						Depth	Reads
<i>Anelloviridae</i>	Torque teno virus AZ1_2	MW679003	3770	52.8	●	59	1466
	Torque teno virus AZ6_5	MW679005	3684	52.1	●	715	17483
	Torque teno virus AZ7_4	MW679006	3781	53.3	●	319	7975
<i>Genomoviridae</i>	Genomovirus AZ4_17	MW679004	2280	53.3	●	24	365
<i>Papillomaviridae</i>	Human papillomavirus AZ1_1	MW679002	7300	36.1	●	39	1908

Influenza A positive
 Influenza B positive
 Influenza A / B negative

B

Anelloviruses

ORF1: capsid protein
 ORF2: regulatory protein
 ORF3: unknown



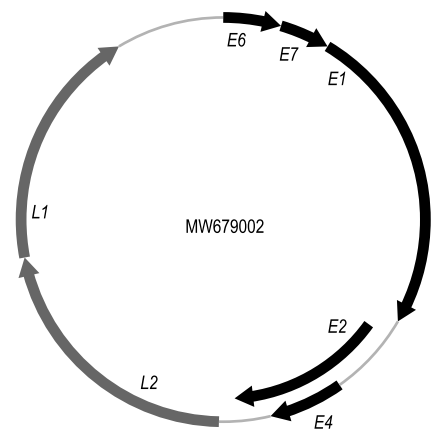
Genomovirus

Cp: capsid protein
 Rep: replication associated protein (initiates replication)



Papillomavirus

E1: DNA replication (helicase)
 E2: transcriptional regulation, initiation of DNA replication and partitioning the viral genome
 E4: contributes to success of genome amplification, virus synthesis & virus release
 E6: oncoprotein binds p53
 E7: oncoprotein binds pRb
 L1: major capsid protein
 L2: minor capsid protein



C

Blastn to hit summary

Family	Genus	Species	Virus	Accession #	Virus	Accession #	Query coverage	E-value	% identity
<i>Anelloviridae</i>	<i>Alphatorquevirus</i>	<i>Torque teno virus 24</i>	Torque teno virus AZ1_2	MW679003	Anelloviridae sp. ctbe000	MH649179	98%	0	86.58%
	<i>Alphatorquevirus</i>	<i>Torque teno virus 29</i>	Torque teno virus AZ6_5	MW679005	Anelloviridae sp. SP31_C2	MZ286093	98%	0	83.01%
	<i>Alphatorquevirus</i>	<i>Torque teno virus 19</i>	Torque teno virus AZ7_4	MW679006	Torque teno virus SAfiA-776-72	MN766622	97%	0	95.7%
<i>Genomoviridae</i>	<i>Gemykolovirus</i>	<i>Gemykolovirus gopha2</i>	Genomovirus AZ4_17	MW679004	Tortoise genomovirus 13	MK570213	99%	0	89.69%
<i>Papillomaviridae</i>	unclassified	unclassified	Human papillomavirus AZ1_1	MW679002	Human papillomavirus mSK_175	MH777317	97%	0	97.41%

FIG 1 (A) Summary of the genomes determined from the four nasopharyngeal swabs with accession number, genome length, GC content, read depth, and coverage. (B) Genome organization of the anelloviruses, genomovirus, and papillomavirus identified from the swab samples. (C) Taxonomic classification of the anelloviruses, genomovirus, and papillomavirus and the summary of the BLASTn analysis.

Sclerotinia sclerotiorum (11) and *Fusarium graminearum* (12). Genomovirus AZ4_17 (GenBank accession number [MW679004](#); 2.28 kb with GC content of 53.3%) shares ~89.6% genome-wide nucleotide identity with tortoise genomovirus 13 ([MK570213](#)) (Fig. 1C) (13) and is part of the genus *Gemykolovirus* and species *Gemykolovirus gopha2* based on the classification guideline for *Genomoviridae* (10). Genomoviruses have been identified previously in human blood, cerebrospinal fluid, and pericardial fluid (14–18), and we speculate that the one identified here may be those that infect oral fungi (yeast).

Papillomaviruses are circular double-stranded DNA viruses (~5.7 to 8.6 kb) and are classified within two subfamilies (*Firstpapillomavirinae* and *Secondpapillomavirinae*) (19). Members of the *Firstpapillomavirinae* encode at least seven genes (19). Human papillomavirus AZ1_1 (MW679002; 7.3 kb with GC content of 36.1%) identified in the swab shares ~97% nucleotide identity with HPV-mSK_175 (MH777317) from a human skin swab (20) (Fig. 1) and HPV_SD2R (KC113191) from an oral swab (21); neither have been assigned a type or classified.

Sample collection was part of routine clinical care which is approved by Arizona State University Institutional Review under study identifier (ID) number STUDY00008985.

Data availability. The sequences of the viruses identified in this study have been deposited in GenBank with accession numbers MW679002, MW679003, MW679004, MW679005, and MW679006. Raw reads have been deposited in SRA project number PRJNA701833 with SRA accessions SRR13720055, SRR13720056, SRR13720057, and SRR13720061.

ACKNOWLEDGMENTS

The research reported in this publication was supported by the National Library of Medicine of the National Institutes of Health under award number U01LM013129 to R.U.H., M.S., and A.V. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

R.U.H. is a cofounder of AquaVitas, LLC, Scottsdale, AZ, an ASU start-up company providing commercial services in wastewater-based epidemiology. R.U.H. is the founder of OneWaterOneHealth, a nonprofit project of the Arizona State University Foundation.

REFERENCES

- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- 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>.
- 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](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Bushnell B. 2014. BMAP: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab., Berkeley, CA.
- Biagini P, Bendinelli M, Hino S, Kakkola L, Mankertz A, Niel C, Okamoto H, Raidal S, Teo CG, Todd D. 2012. Family Anelloviridae, pp 331–341. *In* King AMQ, Adams EB, Carstens EB, Lefkowitz EJ. (ed) *Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses*. Academic Press, Waltham, MA.
- Freer G, Maggi F, Pifferi M, Di Cicco ME, Peroni DG, Pistello M. 2018. The virome and its major component, anellovirus, a convoluted system molding human immune defenses and possibly affecting the development of asthma and respiratory diseases in childhood. *Front Microbiol* 9:686. <https://doi.org/10.3389/fmicb.2018.00686>.
- Varsani A, Opriessnig T, Celer V, Maggi F, Okamoto H, Blomstrom AL, Cadar D, Harrach B, Biagini P, Kraberger S. 2021. Taxonomic update for mammalian anelloviruses (family Anelloviridae). *Arch Virol* 166:2943–2953. <https://doi.org/10.1007/s00705-021-05192-x>.
- 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>.
- Varsani A, Krupovic M. 2017. Sequence-based taxonomic framework for the classification of uncultured single-stranded DNA viruses of the family Genomoviridae. *Virus Evol* 3:vev037. <https://doi.org/10.1093/ve/vev037>.
- Varsani A, Krupovic M. 2021. Family Genomoviridae: 2021 taxonomy update. *Arch Virol* 166:2911–2926. <https://doi.org/10.1007/s00705-021-05183-y>.
- Yu X, Li B, Fu Y, Jiang D, Ghabrial SA, Li G, Peng Y, Xie J, Cheng J, Huang J, Yi X. 2010. A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proc Natl Acad Sci U S A* 107:8387–8392. <https://doi.org/10.1073/pnas.0913535107>.
- Li P, Wang S, Zhang L, Qiu D, Zhou X, Guo L. 2020. A tripartite ssDNA mycovirus from a plant pathogenic fungus is infectious as cloned DNA and purified virions. *Sci Adv* 6:eaay9634. <https://doi.org/10.1126/sciadv.aay9634>.
- Orton JP, Morales M, Fontenele RS, Schmidlin K, Kraberger S, Leavitt DJ, Webster TH, Wilson MA, Kusumi K, Dolby GA, Varsani A. 2020. Virus discovery in desert tortoise fecal samples: novel circular single-stranded DNA viruses. *Viruses* 12:143. <https://doi.org/10.3390/v12020143>.
- Halary S, Duraisamy R, Fancello L, Monteil-Bouchard S, Jardot P, Biagini P, Gouriet F, Raoult D, Desnues C. 2016. Novel single-stranded DNA circular viruses in pericardial fluid of patient with recurrent pericarditis. *Emerg Infect Dis* 22:1839–1841. <https://doi.org/10.3201/eid2210.160052>.
- Lamberto I, Gunst K, Muller H, Zur Hausen H, de Villiers EM. 2014. Mycovirus-like DNA virus sequences from cattle serum and human brain and serum samples from multiple sclerosis patients. *Genome Announc* 2:e00848-14. <https://doi.org/10.1128/genomeA.00848-14>.
- Uch R, Fournier PE, Robert C, Blanc-Tailleux C, Galicher V, Barre R, Jordier F, de Micco P, Raoult D, Biagini P. 2015. Divergent Gemycircularvirus in HIV-positive blood, France. *Emerg Infect Dis* 21:2096–2098. <https://doi.org/10.3201/eid2111.150486>.
- Zhang W, Li L, Deng X, Blumel J, Nubling 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>.
- Zhou C, Zhang S, Gong Q, Hao A. 2015. A novel gemycircularvirus in an unexplained case of child encephalitis. *Virol J* 12:197. <https://doi.org/10.1186/s12985-015-0431-0>.
- Van Doorslaer K, Chen Z, Bernard HU, 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://doi.org/10.1099/jgv.0.001105>.
- Tirosh O, Conlan S, Deming C, Lee-Lin SQ, Huang X, Program NCS, Su HC, Freeman AF, Segre JA, Kong HH, NISC Comparative Sequencing Program. 2018. Expanded skin virome in DOCK8-deficient patients. *Nat Med* 24:1815–1821. <https://doi.org/10.1038/s41591-018-0211-7>.
- Mokili JL, Dutilh BE, Lim YW, Schneider BS, Taylor T, Haynes MR, Metzgar D, Myers CA, Blair PJ, Nosrat B, Wolfe ND, Rohwer F. 2013. Identification of a novel human papillomavirus by metagenomic analysis of samples from patients with febrile respiratory illness. *PLoS One* 8:e58404. <https://doi.org/10.1371/journal.pone.0058404>.