Archives of Virology
https://doi.org/10.1007/s00705-022-05588-3

BRIEF REPORT

A novel statovirus identified in fecal samples from wild geladas in the Ethiopian highlands

India A. Schneider-Crease[1,2,3] · Kenneth L. Chiou[1,2] · Isabella L. Moya[1,2] · Alemayehu Lemma[4] · Ferewihot Ayele[4] · Abebaw Azanaw Haile[5] · Fanuel Kebede[5] · Belayneh Abebe[6] · Thore J. Bergman[7,8] · Jacinta C. Beehner[7,9] · Amy Lu[10] · Michael C. Lund[2,11] · Joy M. Custer[11] · Simona Kraberger[11] · Noah Snyder-Mackler[1,2] · Arvind Varsani[1,2,11,12]

Received: 28 January 2022 / Accepted: 2 August 2022

Abstract

We present a novel statovirus in geladas (*Theropithecus gelada*), graminivorous primates endemic to the Ethiopian highlands. Using a high-throughput sequencing approach, we identified contiguous sequences in feces from two adult female geladas in the Simien Mountains National Park, Ethiopia, that share similarities to statoviruses. Our phylogenetic analysis of the whole genome, as well as the RNA-dependent RNA polymerase (RdRp) and capsid protein (CP) amino acid sequences, revealed that the gelada statoviruses cluster with those from other primates (laboratory populations of *Macaca nemestrina* and *Macaca mulatta*). As the first report of statovirus in wild primates, this finding contributes to our understanding of the phylogenetic and geographic distribution of statoviruses and their hosts.

Keywords Statovirus · *Theropithecus gelada* · Ethiopia · Geladas · Nonhuman primates

Mammalian viromes are integral components of global ecosystems that are shaped by and have the capacity to shape host ecology [1]. Contextualizing the healthy virome composition in host ecology provides a richer understanding of how ecological perturbations might give rise to emerging diseases [1, 2]. Nonhuman primates are a particularly interesting group in which to examine the host-virome relationship given their phylogenetic proximity to humans and their wide geographic and ecological distribution. For example, geladas (*Theropithecus gelada*) are closely related to baboons (*Papio* sp.) but diverge significantly in their ecological niche: where baboons are omnivorous and widely distributed across African biomes, geladas are exclusive graminivores (feeding primarily on grasses, roots,

Handling Editor Ana Cristina Bratanich

✉ India A. Schneider-Crease
indiaaugusta@gmail.com

Noah Snyder-Mackler
nsnyderm@asu.edu

Arvind Varsani
arvind.varsani@asu.edu

[1] Center for Evolution and Medicine, Arizona State University, Tempe, Arizona, USA

[2] School of Life Sciences, Arizona State University, Tempe, Arizona, USA

[3] School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona, USA

[4] College of Veterinary Medicine and Agriculture, Addis Ababa University, Debre Zeyit, Ethiopia

[5] Ethiopian Wildlife Conservation Authority, Addis Ababa, Ethiopia

[6] African Wildlife Foundation, Debark, Ethiopia

[7] Department of Psychology, University of Michigan, Ann Arbor, Michigan, USA

[8] Department of Ecology and Evolution, University of Michigan, Ann Arbor, Michigan, USA

[9] Department of Anthropology, University of Michigan, Ann Arbor, Michigan, USA

[10] Department of Anthropology, Stony Brook University, Stony Brook, New York, USA

[11] The Biodesign Center of Fundamental and Applied Microbiomics, Arizona State University, Tempe, Arizona, USA

[12] Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, 7925 Cape Town, South Africa

Published online: 21 October 2022

and tubers) and are endemic to the Ethiopian highlands [3]. This unique ecological niche has positioned geladas to share more elements of their parasitome with other grazing species, including domestic and wild ungulates, rodents, and lagomorphs, than with closely related baboon species. Indeed, the gastrointestinal helminth parasite community structure of geladas is more similar to that of grazers than to that of baboon species in both species composition and parasite richness [4], and geladas serve as an intermediate host to the tapeworm *Taenia serialis*, which otherwise infects rodents and lagomorphs in its intermediate form [5]. To extend the understanding of how gelada health and disease might differ from that of other primates, we analyzed fecal samples from five wild geladas (three adult females, two adult males) in the Simien Mountains National Park, Ethiopia (13.1510° N, 37.8990° E) for RNA viruses.

Viral RNA was extracted from the five samples using a High Pure Viral Nucleic Acid Kit (Roche Diagnostics, USA). RNA libraries were prepared using an NEBNext Ultra II RNA Library Prep Kit (New England Biolabs) with additional steps for ribosomal and globin depletion following the QIAseq FastSelect RNA removal protocol. Libraries were sequenced on an Illumina Novaseq 6000 S4 flow cell (2 × 150 bp) at the Genomic Sciences Laboratory at North Carolina State University, USA. Adapters and low-quality bases were trimmed using Cutadapt [6], and host reads were removed by mapping to the gelada genome [7]. We then used these non-host reads to assemble the remaining reads using metaSPAdes v 3.15.2 [8]. The *de novo*-assembled contiguous sequences ("contigs") were analyzed by BLASTx [9] against a RefSeq viral protein database (downloaded August 2021) to identify viral-like sequences. We identified three contigs from two samples that were similar to statoviruses: one contig of 4,244 nucleotides (nt) from one adult female (Gel_0447) and two contigs of 1,842 nt and 2,236 nt from another adult female (Gel_0450). While many viral-like sequences were identified in each fecal sample, we report only those similar to statoviruses here.

The 4,244-nt *de novo*-assembled contig (with average depth of coverage of 49.26) from fecal sample Gel_0447 has two open reading frames whose translation products have similarities to RNA-dependent RNA polymerase (RdRp) and capsid protein (CP) sequences (GenBank accession no. OM373193). The contigs (with average depth of coverage of 35.3) from the second fecal sample (Gel_0450) appear to be part of the same genome but are missing a ~ 160-nt sequence in the RdRp coding region. Using a set of specific primers (5'-CTGCATTCAATGTTGAACATAGAGG-3' and 5'-GCAATAACTAAACAATTGCCACTACCAG-3'), we amplified (364-nt amplicon) and cloned this region and sequenced it by the Sanger method. The second statovirus is deposited under GenBank accession no. OM373194 (4238 nt). We also used these specific primers to confirm the presence of statovirus in two of five gelada samples.

Sequences of all statoviruses > 2,500 nt (n = 18) available in GenBank were downloaded (07 Dec 2021). The two statovirus sequences from the gelada samples share 80.4% pairwise nucleotide sequence identity (excluding the spacer) as determined by SDT v1.2 [10], and both share 53.3–68.9% pairwise nucleotide sequence identity with the 18 statovirus sequences downloaded from GenBank (Fig. 1A). This dataset was aligned with MAFFT [11] and used to infer a neighbor-joining phylogenetic tree with 1000 bootstrap replicates. Branches with < 80% bootstrap support were collapsed using TreeGraph2 [12]. The two gelada-feces-derived statoviruses cluster with others from fecal samples of a southern pig-tailed macaque (*Macaca nemestrina*) and three rhesus monkeys (*Macaca mulatta*) (Fig. 1A), forming a nonhuman-primate-specific clade.

We extracted and translated RdRp and CP sequences from the 18 high-coverage statoviruses from GenBank and the two from gelada fecal samples to determine pairwise identity and their phylogenetic relationship. We then aligned the statovirus RdRp and CP amino acid sequences along with representative sequences of tombusviruses (an outgroup including cymbidium ringspot virus, accession no. X15511; pelargonium necrotic spot virus, accession no. J607402), using PROMALS3D [13]. These alignments were used to infer maximum-likelihood phylogenetic trees using IQTREE2 [14] with automatic model selection and aLRT branch support. Branches with < 0.8 aLRT branch support were collapsed using TreeGraph2 [12]. Pairwise amino acid sequence identity values were determined using SDT v1.2 [10]. The RdRp and CP of the two gelada statoviruses share 78.8 and 79.9% amino acid sequence identity, respectively, with each other and 25.3–62.1 and 26.0-55.2% amino acid sequence identity, respectively, with those of other statoviruses. Similar to the genome sequences, the RdRp and CP amino acid sequences cluster with those derived from southern pig-tailed macaques and rhesus macaques (Fig. 1B). Furthermore, the phylogenetic analysis shows two clear clusters, one of statoviruses from primate samples and one from porcine samples (Fig. 1A and B). A single sample from a rhesus macaque falls outside of the primate cluster, with mouse and human samples. Of the available complete or near-complete statovirus sequences, all but one human-feces-derived statovirus from Malawi (KX792981), one wild-mouse-feces-derived statovirus from the United States (KX792983), and the two we discovered in Ethiopian geladas are from animals housed either in farm or laboratory settings in the United States.

Statoviruses are RNA viruses that have been identified in mammalian gastrointestinal tracts and were originally

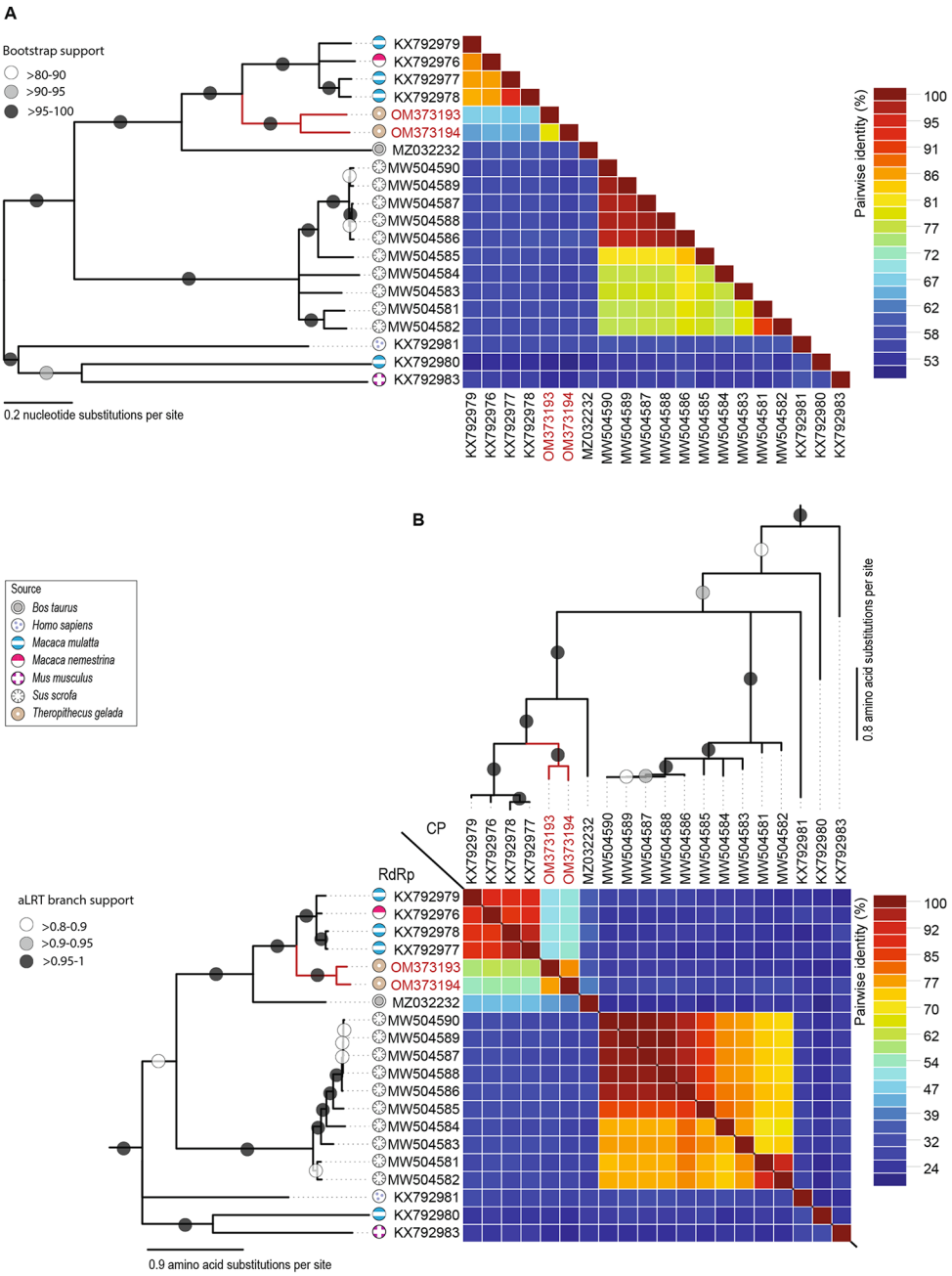

Fig. 1 Phylogenetic and pairwise identity analysis of the statovirus sequences. Branches and accession numbers of the two sequences determined in this study are highlighted in red font and branches. The source of the samples from which statoviruses have been identified is summarized with color-coded unique symbols. (A) Neighbor-joining phylogenetic tree of 18 statovirus (> 2500 nt) nucleotide sequences available in the GenBank database, together with the two derived from gelada samples in this study. A pairwise identity matrix of the nucleotide sequences is provided to the right of the phylogenetic tree. Branches and accession numbers of the two sequences determined in this study are highlighted in red font and branches. The source of the samples from which statoviruses have been identified is summarized with color-coded unique symbols. (B) Maximum-likelihood phylogenetic trees of the RdRp and CP amino acid sequences inferred with the JTT + F + I + G4 and Q.pfam + F + G4 substitution models, respectively, and rooted with sequences of cymbidium ringspot virus and pelargonium necrotic spot virus.

described in metagenomic data from a geographically dispersed set of hosts: children in Malawi, mice, cows, and three populations of captive macaques [15]. They were first identified in 2017 as a monophyletic viral group whose encoded proteins (i.e., RdRp, CP) are currently most closely related to those of tombusviruses (family *Tombusviridae*). Statoviruses have been subsequently identified in metagenomic data from nasal swabs taken from individuals working with animals in Vietnam [16, 17], slurry (fecal and urinary mixture) from a porcine farm (*Sus scrofa*) in the United States [18], and stool from a dairy calf (*Bos taurus*) in the United States [19]. In addition, sera from cows in South Dakota, Pennsylvania, Minnesota, and New York were positive for antibodies that cross-react with recombinantly expressed statovirus capsid protein (expressed in insect cells) [19], and reads mapping to statoviruses were identified in rectal swab samples from semi-domesticated Eurasian tundra reindeer (*Rangifer tarandus*) in Sweden, Norway, and Russia [20]. The broad phylogenetic and geographic spread of statoviruses identified in the four years since their first description suggests that these viruses may be globally distributed across many mammalian species.

Given their taxonomic host breadth, global distribution, and phylogenetic proximity to plant-infecting

tombusviruses and tombus-like virus sequences from invertebrates, it has been proposed that statoviruses may infect components of the mammalian diet rather than the mammalian hosts directly [2]. The detection of statoviruses in human, house mouse, porcine, and bovine samples would thus represent incidental trophic detection (i.e., identification of viruses that infect ingested food items rather than the mammalian host) [21]. Given that geladas are primarily graminivorous in nature, potential hosts might thus include microorganisms (including fungi and protists) or lineages of insects found on plants. It is unlikely that statoviruses infect plants, as we were unable to detect a plant virus movement protein homologous region in these sequences and because they have not yet been identified in plants.

While some statoviruses have been identified in samples from individuals exhibiting clinical symptoms at the time of sampling—such as upper respiratory infections in animal-care workers in Vietnam, enteric disease in a dairy calf in the United States, and fever, cough, and rash in a child in Malawi [15, 16]—these individuals often had concurrent viral and/or parasitic infections that may have produced the observed symptoms. As other isolates were sequenced from asymptomatic individuals, statoviruses may be a component of a healthy virome [2]. Further research should be done to identify the hosts of statoviruses–whether mammalian, protist, arthropod, or other–which will clarify the expectations for associations with symptoms and virome-level interactions in mammals.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s00705-022-05588-3.

Acknowledgements We thank the Ethiopian Wildlife Conservation Authority, the Amhara National Regional Parks State Development and Protection Authority, and all park officials who have permitted and aided our research. We thank Esheti Jejaw, Ambay Fenta, and Setegne Girmay, as well as all past and present field crew of the Simien Mountains Gelada Research Project, for their dedication to over sixteen years of data collection. Long-term gelada research was supported by the National Science Foundation (IOS-1255974, BCS-0715179, BCS-1723237, BCS-1723228), the National Institutes of Health (5R00AG051764-04), the Leakey Foundation, and the National Geographic Society (Gr. #8989-11, Gr. #8100-06). ISC acknowledges funding from the Arizona State University Center for Evolution and Medicine. Finally, we thank the reviewers and editors of this journal for their comments.

Data availability Sequences described in this study have been deposited in the NCBI GenBank database (accession nos. OM373193 and OM373194) and Sequence Read Archive (accession nos. SRR17468237 and SRR17468238).

Statements and declarations

Conflict of interest The authors declare no conflicts of interest.

Compliance with ethical standards All data collection was approved by the Ethiopian Wildlife Conservation Authority and the Arizona State University Institutional Animal Care and Use Committee (approval no. 20-1754R) and conformed to the Code of Best Practices in Field Primatology (American Society of Primatologists).

References

1. French RK, Holmes EC (2020) An Ecosystems Perspective on Virus Evolution and Emergence. Trends Microbiol 28:165–175
2. Koonin EV, Dolja VV, Krupovic M (2021) The healthy human virome: from virus–host symbiosis to disease. Curr Opin Virol 47:86–94. https://doi.org/10.1016/j.coviro.2021.02.002
3. Jarvey JC, Low BS, Pappano DJ, Bergman TJ, Beehner JC (2018) Graminivory and fallback foods: annual diet profile of geladas (*Theropithecus gelada*) living in the Simien Mountains National Park, Ethiopia. Int J Primatol 39:105–126. https://doi.org/10.1007/s10764-018-0018-x
4. Schneider-Crease I, Beehner JC, Bergman TJ, Gomery MA, Koklic L, Lu A et al (2020) Ecology eclipses phylogeny as a major driver of nematode parasite community structure in a graminivorous primate. Funct Ecol 34:1898–1906. https://doi.org/10.1111/1365-2435.13603
5. Schneider-Crease I, Griffin RH, Gomery MA, Dorny P, Noh JC, Handali S et al (2017) Identifying wildlife reservoirs of neglected taeniid tapeworms: non-invasive diagnosis of endemic *Taenia serialis* infection in a wild primate population. PLoS Negl Trop Dis 11:e0005709. https://doi.org/10.1371/journal.pntd.0005709
6. 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
7. Chiou KL, Janiak MC, Schneider-Crease IA, Sen S, Ayele F, Chuma IS et al (2022) Genomic signatures of high-altitude adaptation and chromosomal polymorphism in geladas. Nat Ecol Evol 6:630–643. https://doi.org/10.1038/s41559-022-01703-4
8. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA (2017) metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834. https://doi.org/10.1101/gr.213959.116
9. 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
10. Muhire BM, Varsani A, Martin DP (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PLoS ONE 9:e108277. https://doi.org/10.1371/journal.pone.0108277
11. 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
12. Stöver BC, Müller KF (2010) TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics 11:7. https://doi.org/10.1186/1471-2105-11-7
13. Pei J, Kim B-H, Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res 36:2295–2300. https://doi.org/10.1093/nar/gkn072
14. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A et al (2020) Corrigendum to: IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol 37:2461. https://doi.org/10.1093/molbev/msaa131
15. Janowski AB, Krishnamurthy SR, Lim ES, Zhao G, Brenchley JM, Barouch DH et al (2017) Statoviruses, A novel taxon of RNA viruses present in the gastrointestinal tracts of diverse mammals. Virology 504:36–44. https://doi.org/10.1016/j.virol.2017.01.010

16. Nguyen TTK (2021) Viruses in the upper respiratory tract of individuals at risk of zoonotic infection and their animals in Vietnam: follow-up and virus discovery. Available: https://helda.helsinki.fi/handle/10138/327689
17. Thi Kha Tu N, Thi Thu Hong N, Thi Han Ny N, My Phuc T, Thi Thanh Tam P, van Doorn HR et al (2020) The Virome of Acute Respiratory Diseases in Individuals at Risk of Zoonotic Infections. Viruses 12. https://doi.org/10.3390/v12090960
18. Ramesh A, Bailey ES, Ahyong V, Langelier C, Phelps M, Neff N et al (2021) Metagenomic characterization of swine slurry in a North American swine farm operation. Sci Rep 11:16994. https://doi.org/10.1038/s41598-021-95804-y
19. Hause BM, Nelson E, Christopher-Hennings J (2021) Identification of a novel statovirus in a faecal sample from a calf with enteric disease. J Gen Virol 102. https://doi.org/10.1099/jgv.0.001655
20. Sánchez Romano J, Omazic A, Leijon M, Hagström Ã, Tryland M, Kantanen J et al (2021) Screening of Eurasian Tundra Reindeer for Viral Sequences by Next-Generation Sequencing. Int J Environ Res Public Health 18. https://doi.org/10.3390/ijerph18126561
21. Balique F, Lecoq H, Raoult D, Colson P (2015) Can plant viruses cross the kingdom border and be pathogenic to humans? Viruses 7:2074–2098. https://doi.org/10.3390/v7042074