

8 | Virology | Announcement

Metagenome-assembled microvirus and cressdnavirus genomes from fecal samples of house mice (*Mus musculus*)

Elise N. Paietta,¹ Simona Kraberger,² Joy M. Custer,² Karla L. Vargas,² Erin Ehmke,³ Anne D. Yoder,¹ Arvind Varsani^{2,4}

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT House mice, *Mus musculus*, are highly adapted to anthropogenic spaces. Fecal samples were collected from house mice entering primate enclosure areas at the Duke Lemur Center (Durham, NC, USA). We identified 14 cressdnavirus and 59 microvirus genomes in these mouse feces.

KEYWORDS microvirus, cressdnavirus, *Mus musculus*

nvasive rodents alter ecosystem dynamics by outcompeting endemic wildlife for resources and frequently moving between wild and human-inhabited spaces (1–5). House mice (*Mus musculus*), in particular, have become one of the most widespread invasive species, being transported across the planet through human movement and adapting remarkably well to human-altered landscapes (6).

As part of a larger study sampling lemurs, humans, and rodents at the Duke Lemur Center (Durham, NC, USA) under IACUC #A161-21-08, fecal samples were collected in August 2021 from two M. musculus trapped indoors in lemur enclosure areas. Fecal samples were frozen at -80°C until extraction. Once thawed, fecal samples were homogenized with SM buffer [0.1 M NaCl, 50 mM Tris-HCl (pH 7.4)] and centrifuged briefly at 8,000 rpm. Using the Roche High Pure Viral Nucleic Acid Kit (Roche Diagnostics, Germany), DNA was extracted from 200 µL of each supernatant. DNA extract was amplified using the Templiphi kit (GE Healthcare, USA). After library preparation with the Illumina DNA Prep Kit, libraries were sequenced on an Illumina NovaSeq 6000 at the Duke Center for Genomic and Computational Biology yielding 18,080,238 and 23,175,508 paired reads for each of the two samples. Paired-end reads were trimmed using Trimmomatic v0.39 (7) and assembled with MEGAHIT v.1.2.9 (8). Circular contigs were identified by terminal redundancy based on a >10 nt repeat. Diamond v2.1.9 (9) BLASTx was used to identify viral-like sequences against a local NCBI RefSeg viral protein database (release 220). Cenote Taker2 v2.1.5 (10) and VIBRANT v1.2.1 (11) were used to annotate viral genomes. Viral genomes with >98% identity were clustered into virus operational taxonomic units (vOTUs) with SDT v1.2 (12) and used as a reference to map reads with BBMap v38.12 (13). Web-based BLASTn was used to determine the similarity of viruses characterized in this study to known virus genomes. All bioinformatics tools were used with default settings.

We identified 14 cressdnavirus and 59 microvirus genomes from feces collected from two *M. musculus* individuals (Duke_8, Duke_15). The *Cressdnaviricota* phylum is composed of single-stranded DNA viruses infecting an array of eukaryotic hosts including plants, animals, fungi, and protists (14). The 14 cressdnavirus genomes encode a capsid and replication-associated protein, range in length from 1,847 to 3,474 nt, and GC content of 34.1%-55.0% (Fig. 1A). PP473106 and PP473146 share 97.5% similarity and fall within the *Smacoviridae* family, a group predicted to infect gut archaea (15). PP473147 is a member of the *Genomoviridae* family, a group of likely fungi-infecting cressdnaviruses (16). The unclassified cressdnavirus genomes (n = 11) identified in this

Editor Jelle Matthijnssens, Katholieke Universiteit Leuven, Leuven, Belgium

Address correspondence to Arvind Varsani, arvind.varsani@asu.edu.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 2 April 2024 **Accepted** 20 June 2024 **Published** 8 July 2024

Copyright © 2024 Paietta et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.



			8.15									
А			Duke_8 Duke_1						Top BLAST	n Hit		
Accession No.	% GC	Read depth	Ind.	1 500 1,00	00 1,500	2.000	2,500	3,000 3,474	Accession no.	% identity	% query cover	
PP473142	40.2	2571	- 1 1	, 0,0 1,0	X -	2,000	2,000		KX259437	82.81	99	
PP473143	34.1	7079	-					- I 	KY487919	72.49	49	
PP473144	46.0	219	-				_	L	MT206525	67.07	13	
PP473105, PP473145	50.3	928			X		-	i.	MK858257	72.77	23	
PP473106	55.0	193					-	1	OM892378	68.71	28	
PP473146	55.0	4034	1-1		~		•	1	OM892378 MW436631	69.42 96.93	27 99	
PP473147 PP473148	49.7 46.9	698 7647]		<u>_</u>	1		1	OM154478	71.55	99 15	
PP473149	49.0	499	-		\leftarrow			i.	QDJ95231**	96.00	46	
PP473107	44.2	565	- 11		\rightarrow	_		1	KY487979	98.83	100	
PP473150	55.5	17650	-		\times	-		1	MT206817	72.38	9	
PP473151	39.0	887	-		<u> </u>			i.	KM598399	80.00	4	
PP473152	38.8	5220	-			- !		1	MH617267	84.62	100	
50-100% coverage replication-associated protein												
()-50%	coverage			cal protein							
D			15 8									
В	0/	Deed	Duke 8						Top BLAST			
Accession No.	őc	Read depth	Πġ.	1,000 2,00	0 3,0,00	4,000	5,000	6,000 6,501	Accession no.	% identity	% query cover	
PP473108	36.9	2233 •	(in the second s	1,0,00 2,0,0				- 0,000 0,001	BK052491	89.13	<1	
PP473080	38.9	2228 .		· · ·					BK051593	85.37	89	
PP473081 PP473082	37.2 38.7	19338 · 2123 ·	H	· × ·			Ļ		BK024658 BK045660	68.62 67.03	80 7	
PP473109	38.7	406		1 1					OM869565	68.78	6	
PP473083	38.7	92042		×	X		X	\rightarrow	MT310384	74.70	99	
PP473086	39.6	55813					· ·	\rightarrow	BK054244	89.62	99	
PP473084 PP473085	46.1 39.8	114 · 28566 ·	18				- (MK496711 BK016621	66.58 84.60	5 100	
PP473087	39.4	13462	Hi					<u> </u>	BK016330	82.82	95	
PP473110	40.4	397 •						\rightarrow	BK054244	89.41	99	
PP473088, PP473111	38.6	2938				<u> </u>			OQ686880	68.09	29	
PP473089 PP473090, PP473112	35.5 39.0	302 · 12709 ·					, p t		OQ709391 BK054244	69.10 85.85	38 97	
PP473091	45.3	123		· · · · · · · · · · · · · · · · · · ·					BK030950	79.44	67	
PP473092, PP473113	40.1	68887 ·						\rightarrow	ON482296	72.44	90	
PP473115, PP473114	35.5	325 .	1		×				OP512591	92.16	90	
PP473093 PP473116	43.0 37.5	3793 · 151 ·		· · · · ·	<u> </u>	<u></u>			BK036802 BK016738	67.61 81.76	28 99	
PP473117	36.2	216	H						BK023897	89.38	90	
PP473094, PP473118	42.5	45953 ·	- 1			×			BK046271	69.55	16	
PP473095	30.8 41.9	79 ·							OQ388158 MG883728	70.71 72.84	2 38	
PP473119 PP473096	31.2	122 · 3298 ·		· · · · · · · · · · · · · · · · · · ·					MW678861	76.83	95	
PP473097	36.5	355 -					- X		MT310280	69.00	21	
PP473120	43.7	1044 ·							OR407738	67.07	12	
PP473121 PP473122	34.7 46.4	27 · 1468 ·	H			•			MH617122 OQ686920	66.55 75.52	21 62	
PP473122 PP473098	39.1	3591	l d						OL617036	74.05	8	
PP473123, PP473099	38.1	1622 ·			— ×				ON840047	70.82	35	
PP473124	45.6	460	1				· · — —	į.	OR412218	71.43	3	
PP473100 PP473125	45.3 48.1	43 · 3855 ·						1	OR412218 OQ388114	71.98 70.20	3 30	
PP473125 PP473126	38.7	580	H I i					1	KT264772	86.77	99	
PP473127	39.3	79 ·	-				<u> </u>		MT202774	70.91	23	
PP473128 PP473101, PP473129	42.2	1015 · 117887 ·	- 1						OP884055 OQ686956	69.79 73.02	15 9	
PP473101, PP473129 PP473102, PP473130	41.1 37.4								MW149091	70.44	9 77	
PP473103	41.5	568	- Di		-00				KM589515	87.60	100	
PP473131	53.2	242 ·				*			MG883726	75.90	29	
PP473132	37.3 39.9	874 · 986 ·	Η.					1	OL617014 MH552487	69.16 71.43	19 4	
PP473133 PP473134	39.9 48.1	563	H	· · · ·					MT202150	76.47	41	
PP473135	48.8	379 ·	- 1				• [1	MH616876	75.29	46	
PP473136	43.0	810	1						MT310062	68.52	18	
PP473137 PP473138	43.7 52.1	2124 · 136 ·	H					1	OP884047 MW697604	73.37 69.07	6 27	
PP473138 PP473139	39.9	202	H				1	1	OM154830	67.94	30	
PP473104	38.2	29 -	t 🛛 🕯				1	į.	OM154915	72.66	95	
PP473140	48.2	344 .	1						MW697631	71.91	41	
PP473141	59.0	79					1	1	MT310083	69.03	68	
50-100% coverage Major capsid protein peptidase												
0-50% coverage												
					ural proteir		<u> </u>	ypometical	JULEIN			

FIG 1 (A). Genome organizations of the cressdnaviruses were identified in *M. musculus* feces in this study. A summary of the accession numbers, GC content, read depth, and top BLASTn hit for each distinct genome is presented. **This BLAST hit represents a Rep protein BLASTp instead of a full genome comparison for PP473149. (B) Genome organizations of the microviruses identified in *M. musculus* feces in this study. A summary of the accession numbers, GC content, read depth, and top BLASTn hit for each distinct genome is presented. Genomes characterized in this study with >98% nt identity are represented by one row. For (A) and (B), genome coverage plots based on vOTU read mapping depict the presence of all vOTUs across Duke_8 and Duke_15. Black squares represent 50%–100% genome coverage and serve as a high-confidence proxy of vOTU presence.

study share similarities with viruses identified from water (n = 5), soil (n = 1), fish tissue (n = 1), damselflies (n = 1), capybara feces (n = 1), and tortoise feces (n = 2).

Microviruses are small, single-stranded DNA bacteriophages (17). The 59 microvirus genomes identified encode at least a major capsid protein, DNA pilot protein, and replication initiator protein, range in length from 4,066 to 6,501 nt, and range in GC content from 30.8%–59.0% (Fig. 1B). The distinct (<98% similarity) microvirus genomes (n = 51) share the highest similarity with microviruses characterized from rodent (*M. musculus, Dipodomys merriami, Rattus norvegicus,* and *Sigmodon arizonae*) feces and tissue (n = 8), human samples (n = 17), water (n = 9), ungulate feces (n = 5), soil (n = 3), fish tissue (n = 2), bat feces (n = 2), insects (n = 2), reptile feces (n = 1), avian feces (n = 1), dog feces (n = 1), and airborne particulate (n = 1). The identified microviruses likely infect enterobacteria within *M. musculus*.

ACKNOWLEDGMENTS

The work described here was supported by TriCEM (Triangle Center for Evolutionary Medicine), Duke Biology, Duke Lemur Center, and Sigma Xi grants awarded to E.N.P. We thank the Duke Lemur Center research team for their help in sample collection.

AUTHOR AFFILIATIONS

¹Department of Biology, Duke University, Durham, North Carolina, USA

²The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, Arizona, USA

³Duke Lemur Center, Durham, North Carolina, USA

⁴Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa

AUTHOR ORCIDs

Arvind Varsani b http://orcid.org/0000-0003-4111-2415

FUNDING

Funder	Grant(s) Author(s)
DU Triangle Comparative and Evolutionary Medicine Center,	Elise N. Paietta
Duke University (TriCEM)	Anne D. Yoder
Sigma Xia (SigmaXiSociety)	Elise N. Paietta

AUTHOR CONTRIBUTIONS

Elise N. Paietta, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing – original draft, Writing – review and editing | Simona Kraberger, Investigation, Methodology, Writing – review and editing | Joy M. Custer, Investigation, Methodology, Writing – review and editing | Karla L. Vargas, Investigation, Methodology, Writing – review and editing | Erin Ehmke, Investigation, Methodology, Resources, Writing – review and editing | Anne D. Yoder, Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review and editing | Arvind Varsani, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The sequences of cressdnaviruses and microviruses in this study have been deposited in the NCBI SRA under SRR28214738 and SRR28214739 and GenBank accession numbers PP473080–PP473152.

REFERENCES

- Pittermannová P, Žákovská A, Váňa P, Marková J, Treml F, Černíková L, Budíková M, Bártová E. 2021. Wild small mammals and ticks in Zoosreservoir of agents with Zoonotic potential? Pathogens 10:777. https:// doi.org/10.3390/pathogens10060777
- Chalkowski K, Lepczyk CA, Zohdy S. 2018. Parasite ecology of invasive species: conceptual framework and new hypotheses. Trends Parasitol 34:655–663. https://doi.org/10.1016/j.pt.2018.05.008
- Williams SH, Che X, Garcia JA, Klena JD, Lee B, Muller D, Ulrich W, Corrigan RM, Nichol S, Jain K, Lipkin WI. 2018. Viral diversity of house mice in New York City. mBio 9:e01354-17. https://doi.org/10.1128/mBio. 01354-17
- Pankovics P, Boros Á, László Z, Szekeres S, Földvári G, Altan E, Delwart E, Reuter G. 2021. Genome characterization, prevalence and tissue distribution of astrovirus, hepevirus and norovirus among wild and laboratory rats (*Rattus norvegicus*) and mice (*Mus musculus*) in Hungary. Infection, Genetics and Evolution 93:104942. https://doi.org/10.1016/j. meegid.2021.104942
- Phan TG, Kapusinszky B, Wang C, Rose RK, Lipton HL, Delwart EL. 2011. The fecal viral flora of wild rodents. PLoS Pathog 7:e1002218. https://doi. org/10.1371/journal.ppat.1002218
- WitmerGW, Pitt WC. 2012. Invasive rodents in the United States: ecology, impacts, and management
- 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
- 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

- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth. 3176
- Tisza MJ, Belford AK, Domínguez-Huerta G, Bolduc B, Buck CB. 2021. Cenote-taker 2 democratizes virus discovery and sequence annotation. Virus Evol 7:veaa100. https://doi.org/10.1093/ve/veaa100
- Kieft K, Zhou Z, Anantharaman K. 2020. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome 8:90. https:/ /doi.org/10.1186/s40168-020-00867-0
- 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
- 13. Bushnell B. 2015. BBMap short-read aligner, and other bioinformatics tools. http://sourceforge.net/projects/bbmap/.
- Krupovic M, Varsani A, Kazlauskas D, Breitbart M, Delwart E, Rosario K, Yutin N, Wolf YI, Harrach B, Zerbini FM, Dolja VV, Kuhn JH, Koonin EV. 2020. Cressdnaviricota: a virus phylum unifying seven families of repencoding viruses with single-stranded, circular DNA genomes. J Virol 94:e00582-20. https://doi.org/10.1128/JVI.00582-20
- Díez-Villaseñor C, Rodriguez-Valera F. 2019. CRISPR analysis suggests that small circular single-stranded DNA Smacoviruses infect Archaea instead of humans. Nat Commun 10:294. https://doi.org/10.1038/ s41467-018-08167-w
- Fiallo-Olivé E, Lett J-M, Martin DP, Roumagnac P, Varsani A, Zerbini FM, Navas-Castillo J. 2021. ICTV virus taxonomy profile: Geminiviridae 2021. J Gen Virol 102:001696. https://doi.org/10.1099/jgv.0.001696
- 17. Cherwa JE, Fane BA. 2011. Microviridae: microviruses and gokushoviruses. In eLShttps://doi.org/10.1002/047001590X