

Coding-complete RNA virus genomes assembled from murine cecal metatranscriptomes

Running title: RNA Viral Genomes from Murine Metatranscriptomes

Joshua M.A. Stough, Andrew J. Beaudoin, Patrick D. Schloss[†]

[†] To whom correspondence should be addressed: pschloss@umich.edu

Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109

Abstract

Efforts to catalog viral diversity in the gut microbiome have largely focused on DNA viruses, while RNA viruses remain understudied. To address this, we screened assemblies of previously published mouse gut metatranscriptomes for the presence of RNA viruses. We identified the coding-complete genomes of an Astrovirus and 5 Mitovirus-like viruses.

The viral fraction of the mammalian gut microbiome forms a crucial component in the relationship between microbe and host. Bacterial viruses serve as an important source of genetic diversity and population control for the microbiota, driving its ecology and evolution (1). Mammalian viruses disrupt the gut environment through infection and the response of the host immune system (2). Bacterial and mammalian viruses make significant contributions to host health and disease. Current efforts to describe the diversity of viruses present in the gut have focused on using shotgun metagenomics to identify double-stranded DNA viruses, predominantly bacteriophage and host pathogens (3). However, this method ignores viruses with RNA genomes, which make up a considerable portion of the environmental viromes (4).

We re-analyzed deeply-sequenced metatranscriptome data produced by our lab for the study of microbiome dynamics in a mouse model of *Clostridioides difficile* infection (5, 6). Briefly, C57Bl/6 mice from a breeding colony we maintain at the University of Michigan were treated with one of three different antibiotics (clindamycin, streptomycin, or cefoperazone). After a 24 hour recovery period, the mice were infected with *C. difficile* strain 630. Germ free C57Bl/6 mice were also mono-associated with *C. difficile* strain 630. Cecal contents were removed from each animal 18 hours post infection and frozen for RNA extraction and sequencing. RNA sequences from each sample were trimmed of adapter sequences and low-quality bases using Trimmomatic v0.39 and assembled individually using rnaSPAdes v3.13.1 (7) and concatenated for dereplication, resulting in 70,779 contigs longer than 1 kb. Contigs were screened for the presence of RdRP coding sequences using Blast v2.9.0 against a database containing all Viral RefSeq protein sequences annotated as RdRP (screening database available online as described below) with a maximum e-value of 10^{-20} , resulting in 29 contigs. RdRP is conserved amongst almost all RNA viruses without a DNA stage in genome replication. These contigs were then annotated with Interproscan v5.39-77.0 (8, 9). We constructed phylogenetic trees from RdRP protein sequences using IQ-TREE v1.6.12 (10).

Two classes of RNA viruses were assembled with high coverage with sequences originating from most of the mouse treatment groups, including germ-free mice. First, a 6,811 base-long astrovirus genome (GC 56.6%) was obtained with 1,683.5-fold coverage (Figure 1A). The genome contained 3 predicted open reading frames encoding a capsid, RdRP, and a trypsin-like peptidase, and appears

to be closely related to Murine Astrovirus in *Astroviridae*. Second, 5 distinct, but closely related RNA virus genomes, designated Putative Mitovirus JS1 through JS5, ranging in length from 2,309 to 2,447 bases with 4.6 to 16,078.8-fold coverage and average GC content of 46.2% belonged to a previously undescribed clade of *Narnaviridae* adjacent to the Mitoviruses (Figure 1B). These RNA virus genomes will facilitate future studies of RNA virus biology in the murine microbiome.

Data Availability. The RNA-seq data are available the NCBI Sequence Read Archive (SRA) database under the accession numbers PRJNA354635 (*C. difficile* infected mice) and PRJNA415307 (mock-infected mice). The assembled genomes are available at the National Center for Biotechnology Information (NCBI) GenBank under the accession numbers MN780842-MN780847. All of the scripts and software used to perform this analysis are available at https://github.com/SchlossLab/Stough_Mouse_RNA_Virome_MRA_2019.

Acknowledgements

This research was supported by NIH grant U01AI12455. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

References

1. **Ogilvie LA, Jones BV.** 2015. The human gut virome: A multifaceted majority. *Frontiers in Microbiology* **6**:918. doi:10.3389/fmicb.2015.00918.
2. **Legoff J, Resche-Rigon M, Bouquet J, Robin M, Naccache SN, Mercier-Delarue S, Federman S, Samayoa E, Rousseau C, Piron P, Kapel N, Simon F, Socié G, Chiu CY.** 2017. The eukaryotic gut virome in hematopoietic stem cell transplantation: New clues in enteric graft-versus-host disease. *Nature Medicine* **23**:1080–1085. doi:10.1038/nm.4380.
3. **Garmaeva S, Sinha T, Kurilshikov A, Fu J, Wijmenga C, Zhernakova A.** 2019. Studying the gut virome in the metagenomic era: Challenges and perspectives. *BMC Biology* **17**:84. doi:10.1186/s12915-019-0704-y.
4. **Culley A.** 2018. New insight into the RNA aquatic virosphere via viromics. *Virus Research* **244**:84–89. doi:10.1016/j.virusres.2017.11.008.
5. **Jenior ML, Leslie JL, Young VB, Schloss PD.** 2017. *Clostridium difficile* colonizes alternative nutrient niches during infection across distinct murine gut microbiomes. *mSystems* **2**:e00063–17. doi:10.1128/mSystems.00063-17.
6. **Jenior ML, Leslie JL, Young VB, Schloss PD.** 2018. *Clostridium difficile* alters the structure and metabolism of distinct cecal microbiomes during initial infection to promote sustained colonization. *mSphere* **3**:e00261–18. doi:10.1128/mSphere.00261-18.
7. **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. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* **19**:455–477. doi:10.1089/cmb.2012.0021.
8. **UFBoot2: Improving the ultrafast bootstrap approximation.**
9. **Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler A von, Jermiin LS.** 2017. ModelFinder:

74 Fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**:587–589.
75 doi:10.1038/nmeth.4285.

76 10. **Nguyen L-T, Schmidt HA, Haeseler A von, Minh BQ.** 2015. IQ-TREE: A fast and effective
77 stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and*
78 *Evolution* **32**:268–274. doi:10.1093/molbev/msu300.

79 **Figure 1. Phylogenetic trees showing the relatives of the metatranscriptome assembled**
80 **genomes.** Maximum Likelihood phylogenetic trees constructed from RdRP amino acid sequences
81 for (A) Astroviruses and (B) Narnaviruses. Node annotations represent IQTree Ultra-fast Bootstrap
82 statistics, values less than 50% were excluded from the tree. Scale bars are marked in red to the left
83 of each tree. Highlight colors in (B) represent major Narnavirus taxa: Orange - Ourmiaviruses, Pink
84 - Ourmia-like Mycoviruses, Gray - Narnaviruses, Blue - Mitoviruses, Purple - Murine Mitovirus-like
85 viruses, Green - Leviviruses.