Mouse reference gut virome catalog that enable comprehensive understanding of mouse gut virosphere

**Han-June Kim**[1,3#], _**Nayeon Kim**_[1#], Jun Hyung Cha[1], Wonjong Kim[1], Junyeong Ma[1], Jungyeon Kim[1], Yerin Kim[1], Martin Hemberg[2], Insuk Lee[1,3*]

[1]Department of Biotechnology, College of Life Science and Biotechnology, Yonsei University, Seoul 03722, Republic of Korea

[2]Gene Lay Institute of Immunology and Inflammation, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

[3]DECODEBIOME Co., Ltd. Incheon 21983, Republic of Korea

[#]These authors contributed equally to this work

[*]Corresponding author: Insuk Lee Department of Biotechnology, College of Life Science and Biotechnology, Yonsei University, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea Tel: +82-10-4186-8706, E-mail: insuklee@yonsei.ac.kr

Abstract

Recent studies have shown that not only gut bacteria but also gut viruses play important roles in human health. While bacterial communities in the human gut microbiome have been extensively studied, viral communities (the virome) have only recently begun to receive attention, despite their likely functional influence on the overall gut ecosystem. This limited understanding hinders a comprehensive view of the human microbiome. Mouse models have proven valuable in elucidating mechanisms of host–bacteria interactions in the gut, and similarly, mice can serve as effective models for studying how viruses impact host physiology. To address this, we developed an end-to-end pipeline to reconstruct high-quality viral metagenome-assembled genomes (vMAGs), including a viral genome binning step. By analyzing 2,638 mouse gut whole metagenome shotgun sequencing (WMS) datasets, we generated the Mouse Reference Gut Virome (MRGV), a comprehensive genomic catalog of the mouse gut virome. MRGV comprises 109,778 near-complete viral genomes, representing 28,824 species-level viral operational taxonomic units (vOTUs). In total, we identified 8,212,891 protein-coding sequences, which were clustered into 625,774 protein families, 47% of which were functionally annotated using 3D structural homology. To our knowledge, the MRGV is the first publicly available resource that provides both taxonomic and functional information on the mouse gut virome. This reference will be instrumental in advancing our understanding of how the gut virome, in both mice and humans, influences host physiology.