Exploratory factor analysis can be a novel standard method for virome analyses

Seiichi Omura, Alfredo A. Hinay Jr., Ikuo Tsunoda

Department of Microbiology, Kindai University Faculty of Medicine, Osaka, Japan

Seiichi Omura, Ph.D.
somura@med.kindai.ac.jp

Abstract

Background: Gut microbiota comprises various microbes, including bacteria and viruses. Although the gut bacteriome has been associated with several diseases, including multiple sclerosis (MS), the gut virome has not been well-characterized in both health and disease conditions. This could be due to the methodology used to determine the microbiome: although principal component/coordinate analysis (PCA/PCoA) is a gold standard method for bacteriome analyses, it has failed to characterize virome profiles. Exploratory factor analysis (EFA) is a statistical method frequently used in the psychological field, as it can identify the latent factors/causes among the variables. Previously, using human gut virome data from patients with inflammatory bowel disease (IBD) (n = 12) and controls (n = 12), we successfully applied EFA to characterize the virome. EFA clearly separated the viruses/bacteriophages (phages) into three groups: Group 1, phages infecting *Klebsiella*, which have been associated with IBD; Group 2, phages infecting enterobacteria; and Group 3, viruses infecting mammalian cells (Omura and Tsunoda, The 4th Annual Meeting of the Japanese Society for Phage Therapy, 2024). In this study, we applied EFA to characterize large gut virome data from the database, which contained larger-sized data (average: 382,271 reads/sample) from a greater number of samples (n = 96) than the IBD data (34,956 reads/sample, n = 24) that we previously used.

Methods: We used shotgun metagenome data from the fecal DNA of patients with MS (n = 45) and controls (n = 51) from a public database (Accession no. PRJNA1084182). We extracted virome data and analyzed them with PCA and EFA.

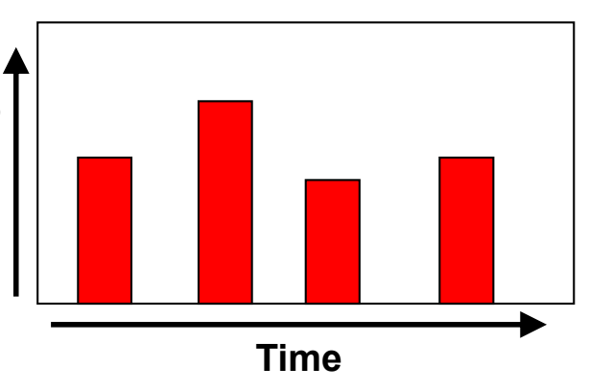
Results: In both the MS and control metagenome data, 1,525 species of DNA viruses/phages were included. Although PCA of virome data was unable to distinguish between patients and controls, EFA extracted six factors from the virome data. Factor 1 contained 269 phages infecting Gram-positive bacteria, including the genus *Propionibacterium*. Factor 2 contained 117 phages infecting Gram-negative bacteria, including the genera *Escherichia* and *Pseudomonas*. Factor 3 contained 332 phages, whose host bacteria included the genera *Enterococcus* and *Salmonella*. All viruses in Factor 4 were 113 Streptococcus phages belonging to the family *Aliceevansviridae*. Factor 5 contained 61 phages infecting the genus *Staphylococcus* and some Gram-negative bacteria. Factor 6 was composed of nine Lactobacillus phages and 36 Lactococcus phages. Comparing the factor scores between the MS and control groups, Factors 1 and 5 were lower and higher in MS patients, respectively, than in controls ($P < 0.01$, chi-squared test). Although the other factors didn't show differences between the groups, Factor 6 may be associated with dairy food consumption or commensal lactic acid bacteria. Unclassified viruses included viruses infecting mammalian cells.

Conclusions: We demonstrated that EFA could be applied to large-sized virome data, classifying the virome into six distinct viral groups based on unique sets of host bacteria that phages can infect. Among the Factors, Factors 1 and 5 were strongly associated with MS; phages in Factors 1 and 5 may regulate unique sets of bacteria, contributing to changes in gut bacteriome, or can be useful as a biomarker of MS. EFA can be used as a novel standard method for virome analyses.

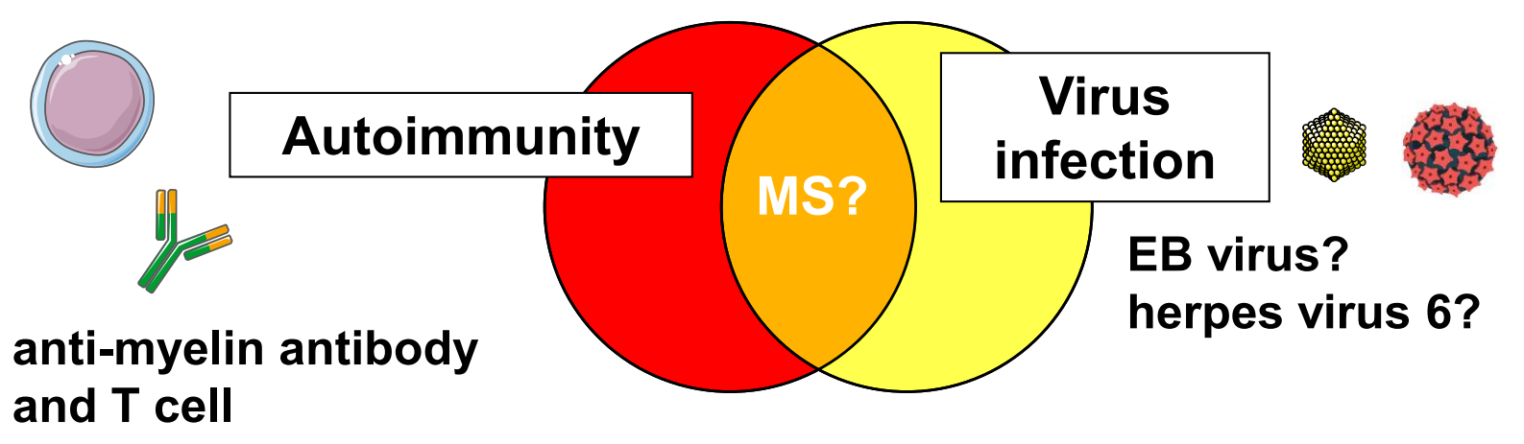
Introduction

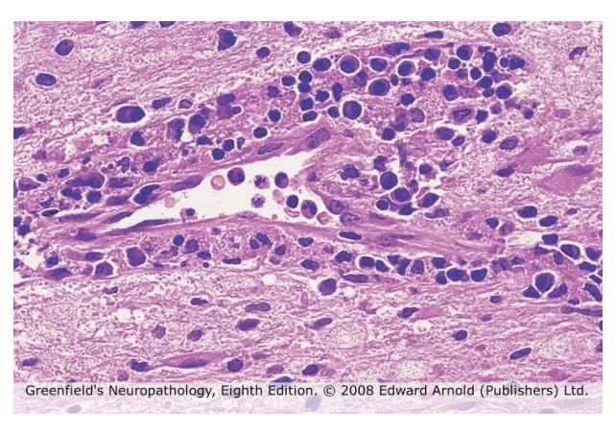
Multiple sclerosis (MS)

- MS is a chronic immune-mediated disease in the brain and spinal cord
- Distinct episode of neurologic attacks, separated in time, attributed to white matter lesions that are separated in space ("multiple")
- Inflammation in the brain or spinal cord
  ➡Gliosis (firm scars, "sclerosis")
- Anti-CNS autoimmune responses and/or virus infections are proposed to damage the nerve fibers, leading to neurological deficits, including motor paralysis



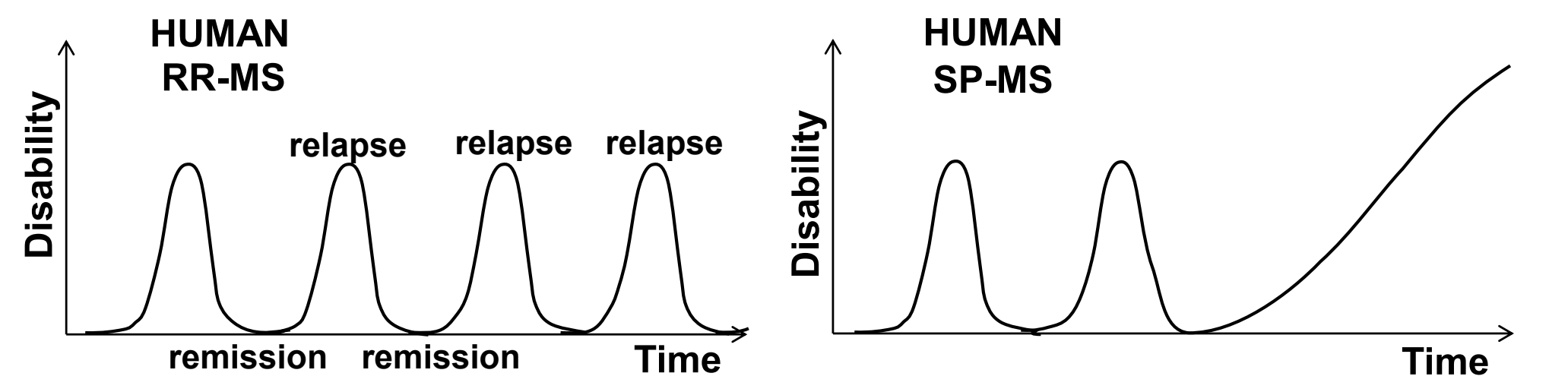
Park A-M, Omura S et al., *Clin Exp Neuroimmunol*, 2017



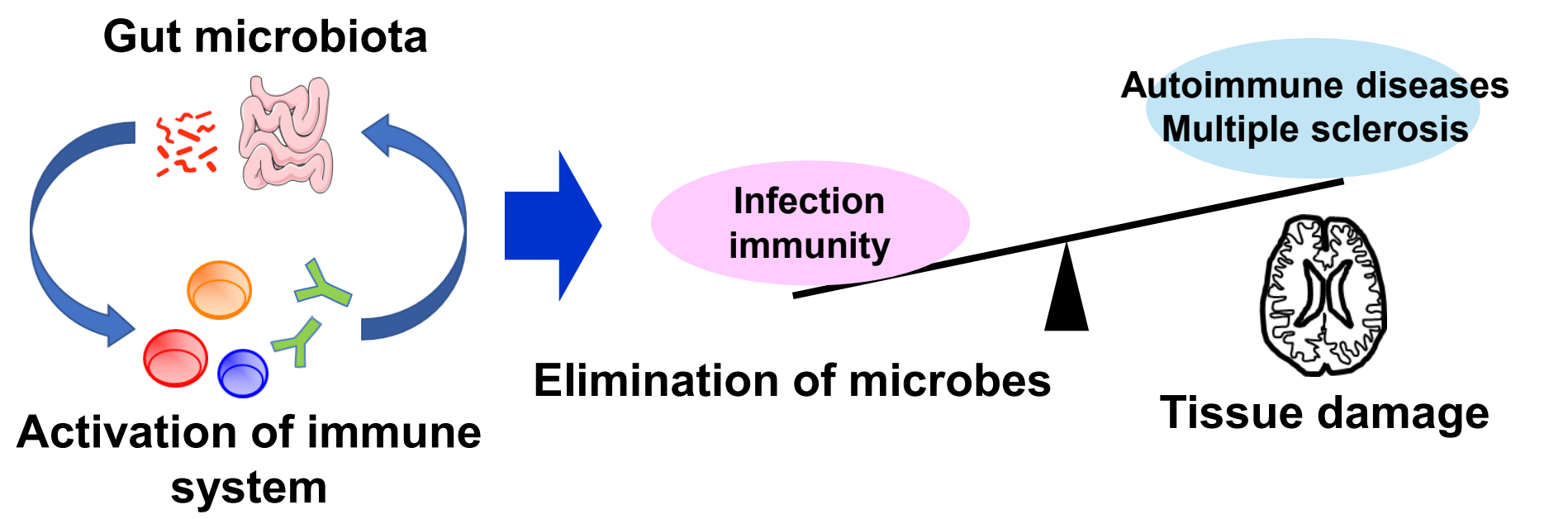


Sato et al., *Neuroinflammation*, 2018

- MS patients show various clinical courses, including relapsing-remitting (RR) MS and secondary progressive (SP) MS



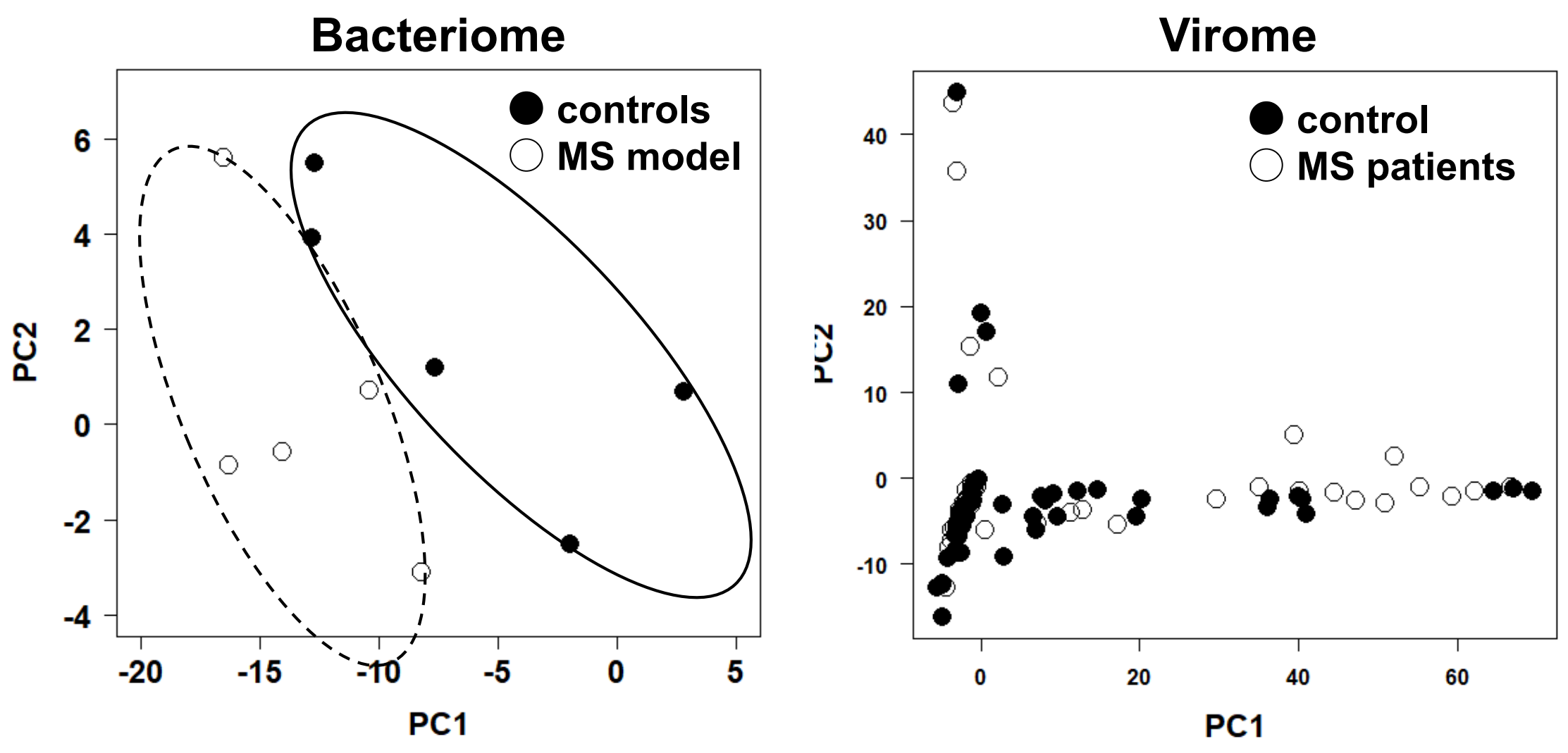
Multiple sclerosis (MS) and microbiome



Park A-M, Omura S et al., *Clin Exp Neuroimmunol*, 2017

- Gut microbiota can activate systemic immune responses contributing to anti-microbial immunity.
- Uncontrolled excessive immune responses cause immune-mediated tissue damage in remote organs.
  ➡autoimmune diseases and/or MS
- Gut bacteriome modulation may suppress MS.
- Role of gut virome in MS has not been clarified.

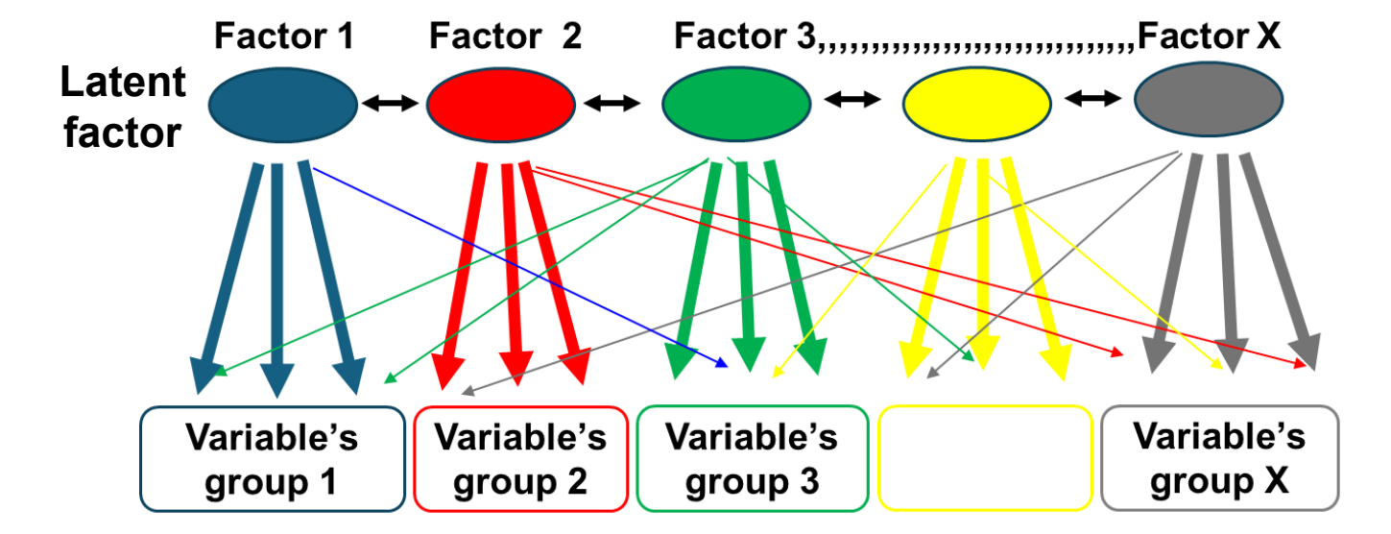
Principal component analysis (PCA) of bacteriome and virome data



- PCA of gut bacteriome data from MS model clearly separated from control mice (Omura et al., 2020).
- PCA of gut virome data did not distinguish between the MS patients and healthy controls.

Factor analysis

- Factor analysis has been used in the psychology field but not in biomedical field.
- Factor analysis can identify the latent factors/causes among the variables.



Aim

"To characterize the virome data of MS and other diseases by factor analysis"

Methods

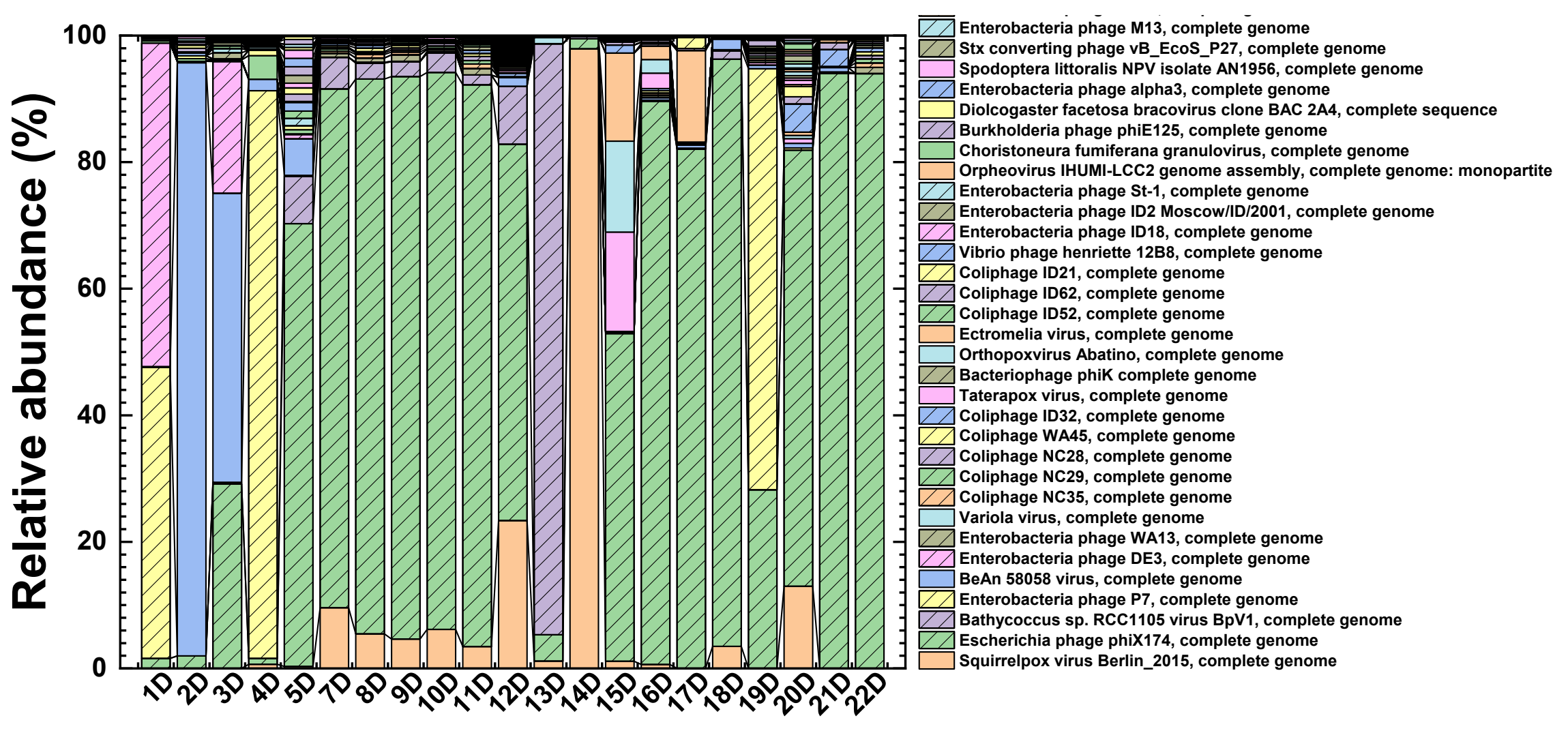
The virome data were obtained as follows:

- Virome analysis of 22 lacrimal fluid samples using MinION device
- Virome data extracted from shotgun metagenome data (PRJNA391511, Fernandes et al., *J Pediatr Gastroenterol Nutr*, 2019) from 12 IBD patients and 12 healthy controls
- Virome data extracted from shotgun metagenome data (PRJNA1084182, Ghimire et al., *Proc Natl Acad Sci USA*, 2025) from 45 MS patients and 51 healthy controls

Virome data analysis
⇒ Mapping to RefSeq viral reference genome and counting read fragments
⇒ Factor analysis
→ Determining factor number by scree plot
→ Exploratory factor analysis by an R package "Psych"
→ Grouping by factor loadings
→ Latent factor identification

Results

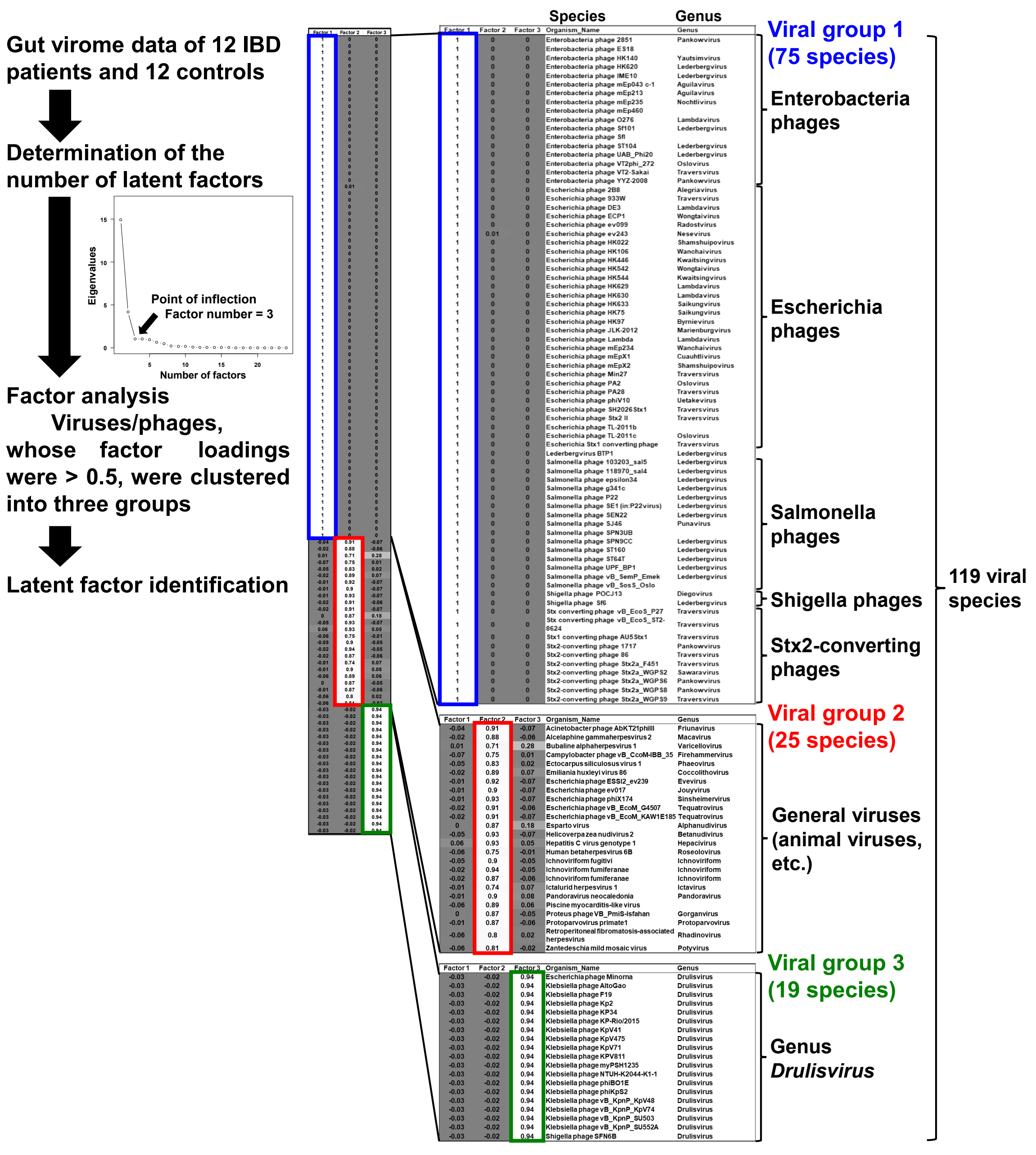
Virome is dominantly composed of bacteriophages (phages)



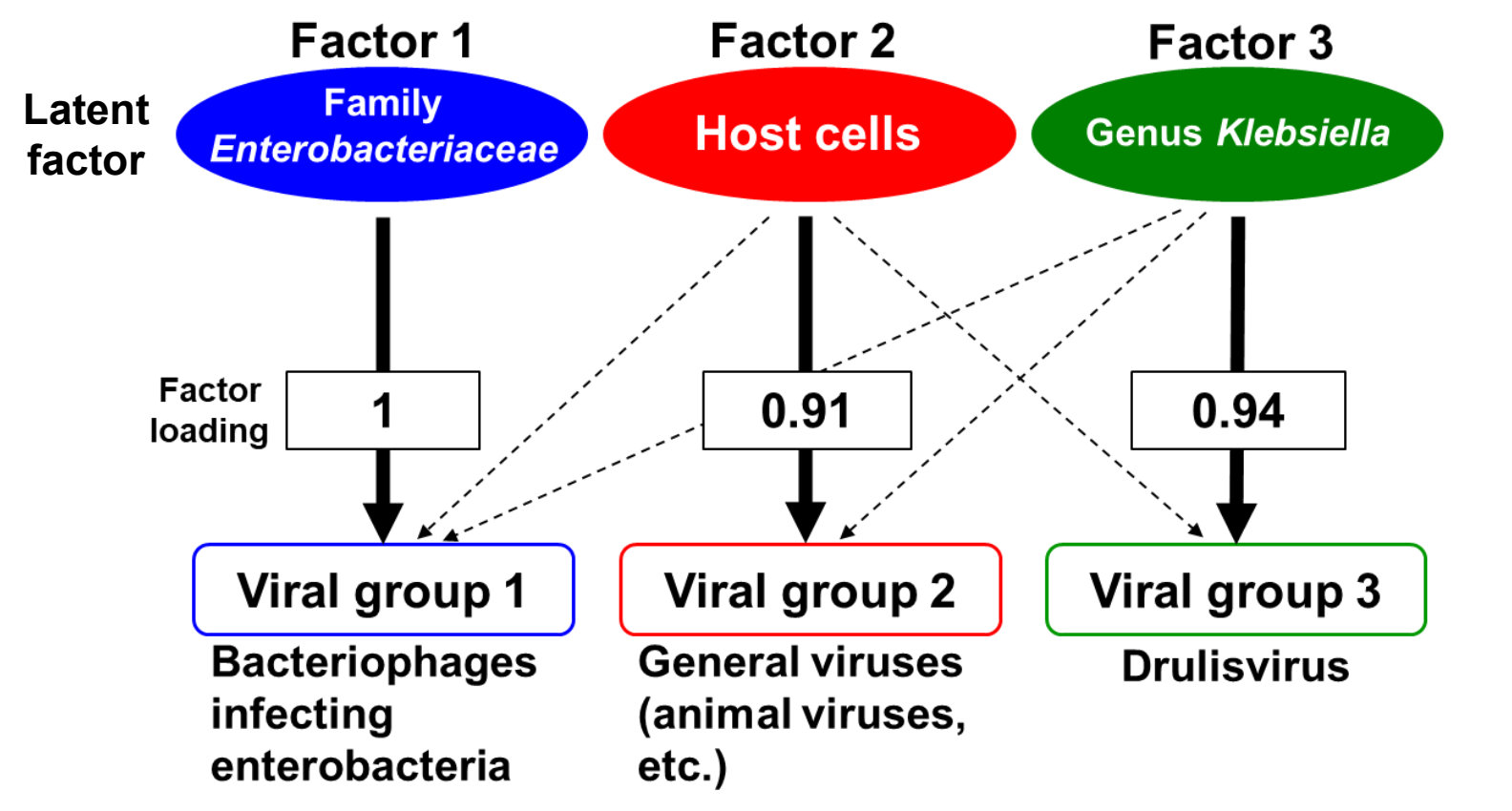
We determined the virome in human lacrimal fluid and found that virome was mostly composed of phages shown in hatched bars.

Factor analysis of gut virome in inflammatory bowel disease (IBD)

- 119 viral species (total 831,496 reads) were identified.



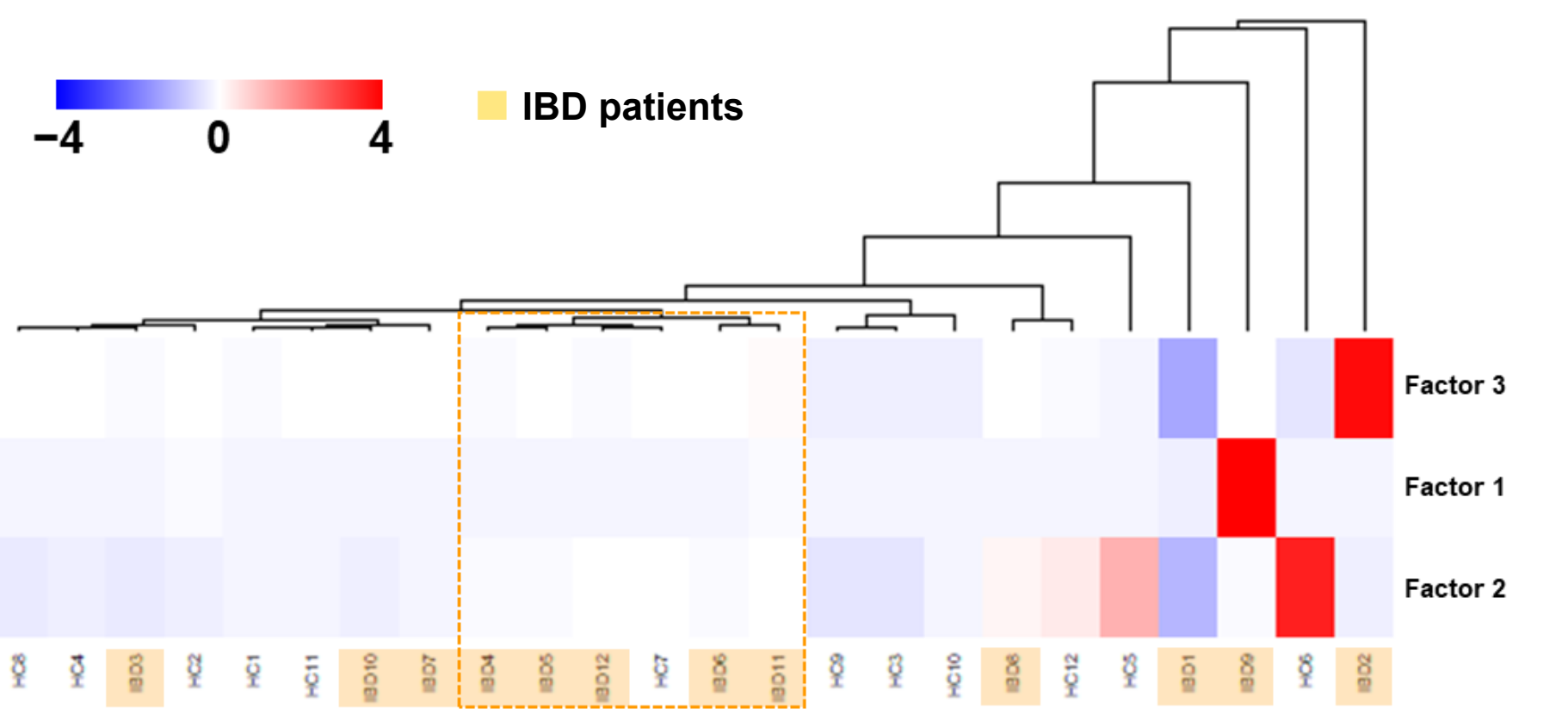
Gut virome data of 12 IBD patients and 12 controls
↓
Determination of the number of latent factors
↓
Factor analysis
Viruses/phages, whose factor loadings were > 0.5, were clustered into three groups
↓
Latent factor identification



- Factor analysis identified three latent factors that were associated with bacteriophages (phages)/viruses infecting distinct hosts.
- Factor 1 appeared to be the phages infecting the Family *Enterobacteriaceae*.
- Factor 2 seemed to be general viruses, including animal viruses.
- Since Factor 3 was strongly associated with some IBD patients, bacteria in the genus *Klebsiella*, which has been linked to IBD etiologically (Federici et al., 2022), seemed pathogenic for some patients.

Dendrogram of IBD patients and controls based on the factor scores

Factor scores indicate the positive (🟥) or negative (🟦) association between the patients and factors.



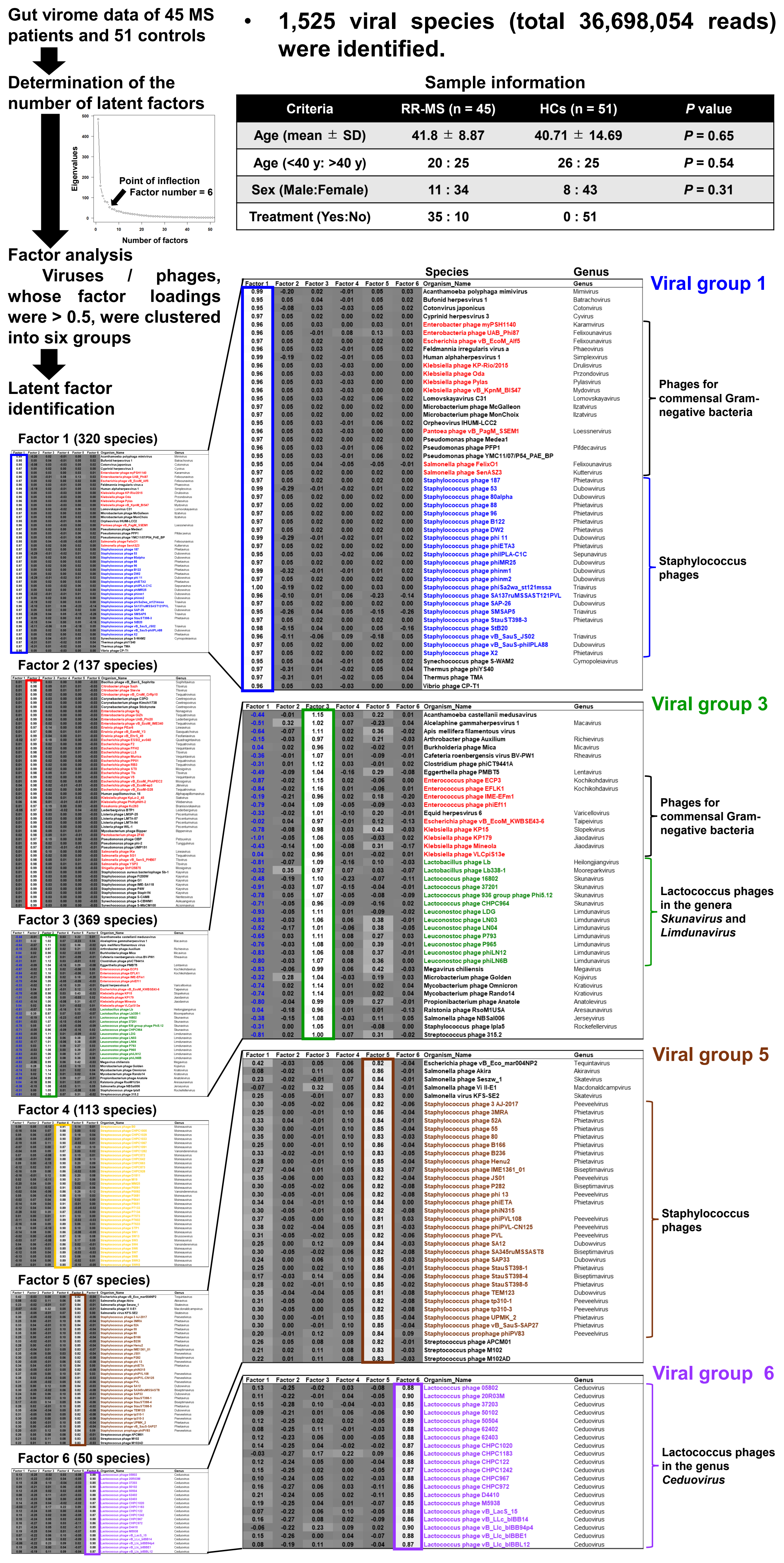
- When a group composed of 3 or more people is defined as "Cluster," a total of 12 samples of IBD and controls (HCs) were divided into three clusters and others. Five of six (83%) of samples in Cluster 2 were from IBD patients.

Factor analysis of gut virome in MS

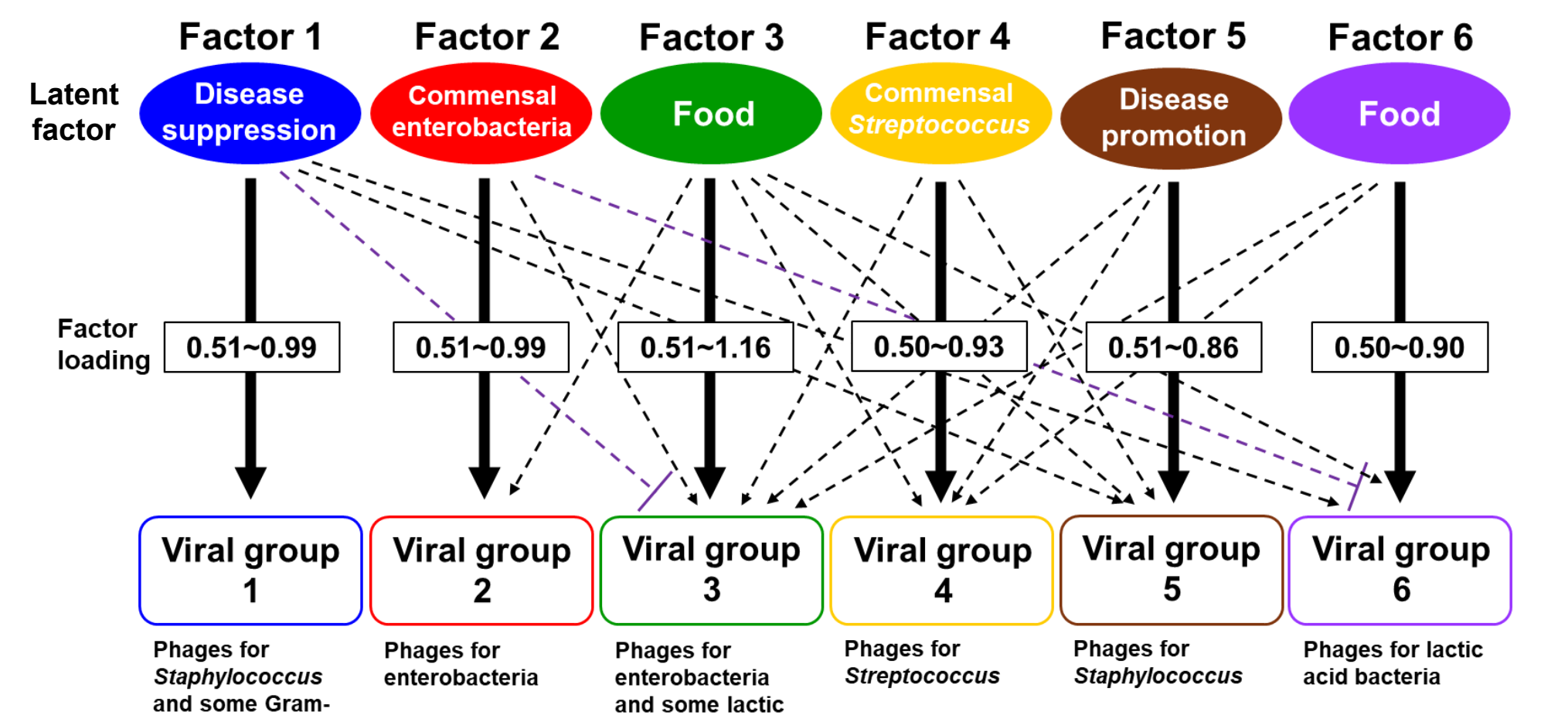
- 1,525 viral species (total 36,698,054 reads) were identified.

Sample information

Criteria	RR-MS (n = 45)	HCs (n = 51)	*P* value
Age (mean ± SD)	41.8 ± 8.87	40.71 ± 14.69	*P* = 0.65
Age (<40 y: >40 y)	20 : 25	26 : 25	*P* = 0.54
Sex (Male:Female)	11 : 34	8 : 43	*P* = 0.31
Treatment (Yes:No)	35 : 10	0 : 51	

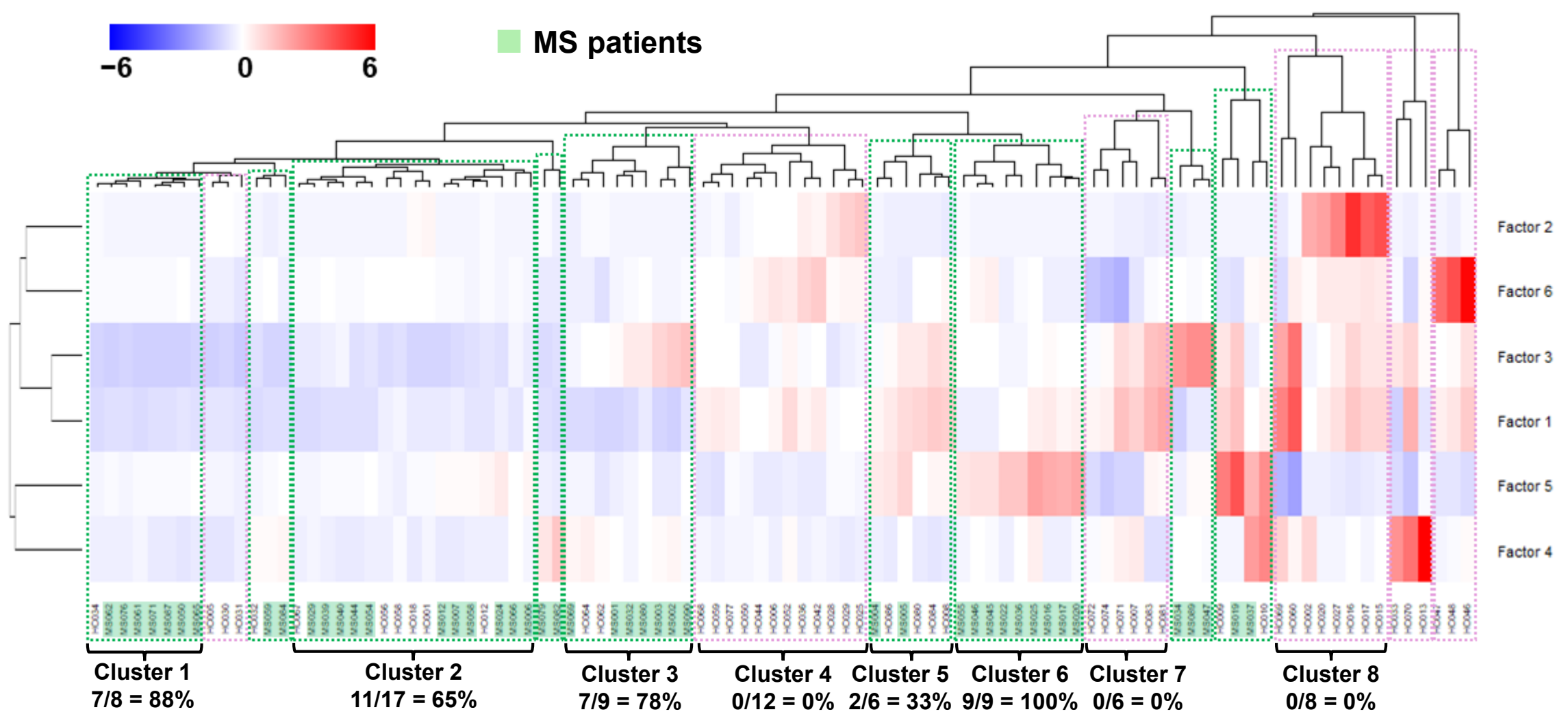


Gut virome data of 45 MS patients and 51 controls
↓
Determination of the number of latent factors
↓
Factor analysis
Viruses / phages, whose factor loadings were > 0.5, were clustered into six groups
↓
Latent factor identification



- Factor scores of factors 1 and 5 were associated with MS patients negatively and positively, which may indicate disease suppression and promotion, respectively (MS vs. factor 1 or 5, $P < 0.01$, chi-square test).
- Factors 2 and 4 appeared to be commensal bacteria; Factors 3 and 6 seemed to be food.
- Factors 1 and 2 might suppress viral groups 3 and 6, respectively.

Dendrogram of MS patients and controls based on the factor scores



- When a group composed of six or more people is defined as "Cluster," a total of 96 samples of MS and HCs were divided into 8 clusters and others. Several clusters included a high proportion of MS patients.
- Clusters 4, 7, and 8 contained no MS patients

Conclusions

- PCA didn't distinguish the virome between MS patients and controls.
- Factor analysis could identify the latent factors that were associated with unique sets of bacteriophages.
  - In IBD, factor analysis identified viral group 3, which contains Klebsiella phages in genus *Drulisvirus*, that might be associated with disease pathogenesis.
  - In MS, bacteriophages in Factors 1 and 5 might contribute to the disease conditions of MS.
  - A dendrogram of factor scores may result in the identification of distinct subpopulations in IBD and MS.

References

- Omura S, et al., (2020). Bioinformatics analysis of gut microbiota and CNS transcriptome in virus-induced acute myelitis and chronic inflammatory demyelination; potential association of distinct bacteria with CNS IgA upregulation. *Front Immunol*, 11:1138. doi: 10.3389/fimmu.2020.01138.
- Omura S, et al. (2022). Exploratory factor analysis determines latent factors in Guillain–Barré syndrome. *Scientific Reports*, 12:21837. doi: 10.1038/s41598-022-26422-5.
- Ahmad I, Omura S, et al. (2025). The safety of *Alcaligenes* lipid A in an immune-mediated disease model associated with IgA, Th17 cells, and gut microbiota. *Int J Mol Sci*. In review.

This study has no conflict of interest.