

Factor analysis of the gut virome determines unique sets of bacteriophages associated with multiple sclerosis

Seichi Omura, Ikuo Tsunoda

Kindai University Faculty of Medicine, Osaka, Japan



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



Abstract

Background: Gut microbiota comprises various microorganisms, 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 health and disease conditions. This limitation may be due to the methodological challenges in profiling the microbiome. Principal component/coordinate analysis (PCA/PCoA), a gold standard for gut bacteriome analyses, has failed to characterize virome profiles. Factor analysis (FA) is a bioinformatics method frequently used in the psychological field, since FA can identify the latent factors/causes among the variables. Previously, we applied FA in biomedical research and classified 55 anti-glycolipid antibodies into five groups in Guillain-Barré syndrome (Omura et al., 2022). Here, we aimed to characterize the gut virome of MS using FA.

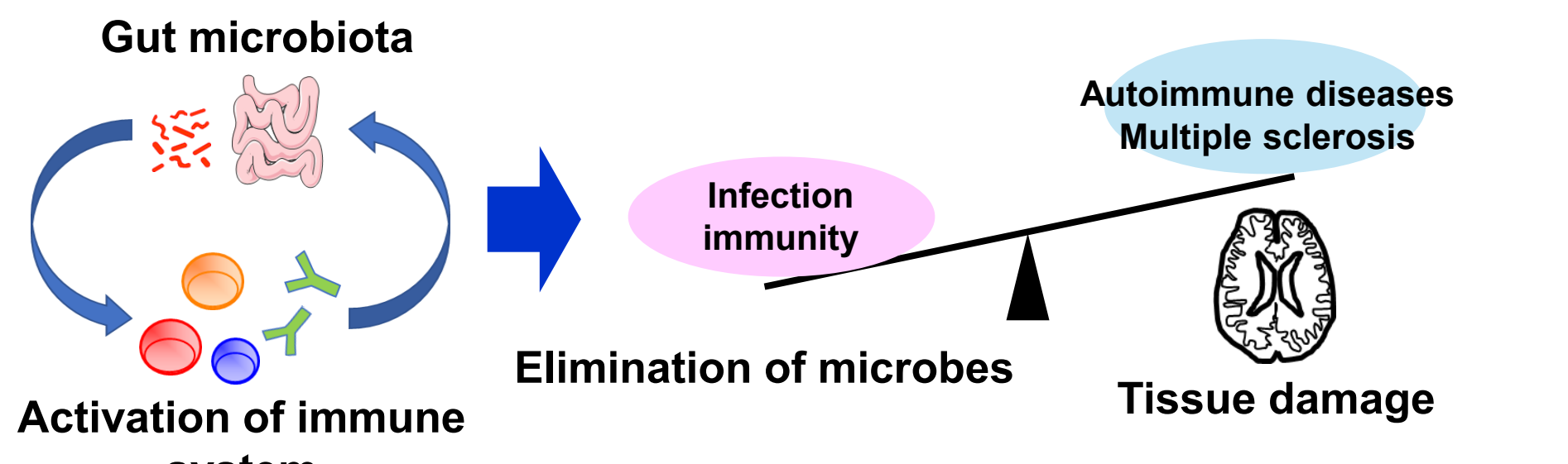
Methods: We used shotgun metagenome data from fecal DNA of the patients with relapsing-remitting MS (n = 45) and controls (n = 51) from the database (Accession no. PRJNA1084182). We extracted virome data and analyzed them with PCA and FA using R.

Results: Using both MS and control metagenome data, we extracted virome data, which contained 1,525 species of DNA viruses/bacteriophages (phages). PCA of virome data could not separate the patients from the controls. On the other hand, FA extracted six factors from the virome data. Factor 1 contained 269 phages infecting Gram-positive bacteria, including the genus *Staphylococcus*. Factor 2 contained 117 phages infecting enterobacteria, including the genera *Escherichia* and *Salmonella*. Factor 3 contained 332 phages, whose host bacteria included lactic acid bacteria and enterobacteria. All viruses in Factor 4 were 113 Streptococcus phages. 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 controls ($P < 0.01$). 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.

Conclusions: We demonstrated that FA could classify gut 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. FA may be useful as a biomarker or to identify the phages regulating unique sets of bacteria, contributing to dysbiosis in MS.

Introduction

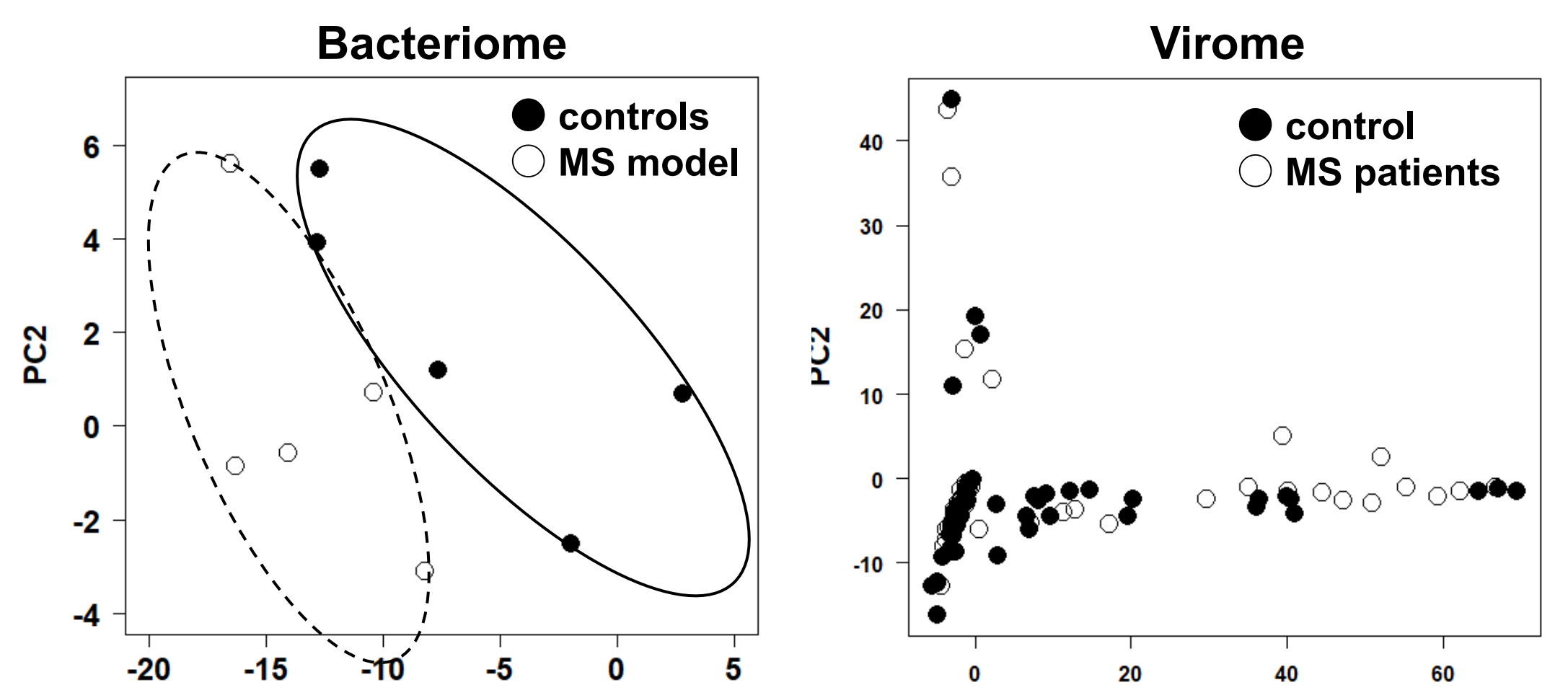
Multiple sclerosis (MS) and microbiome



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.

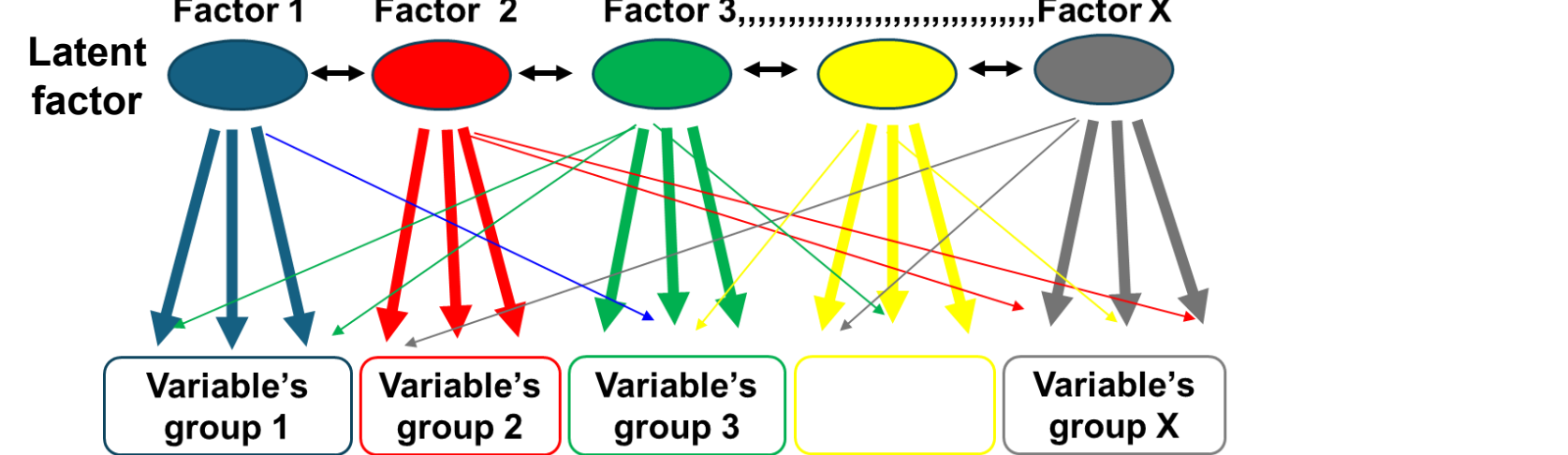
Principal component analysis (PCA) of bacteriome and virome data



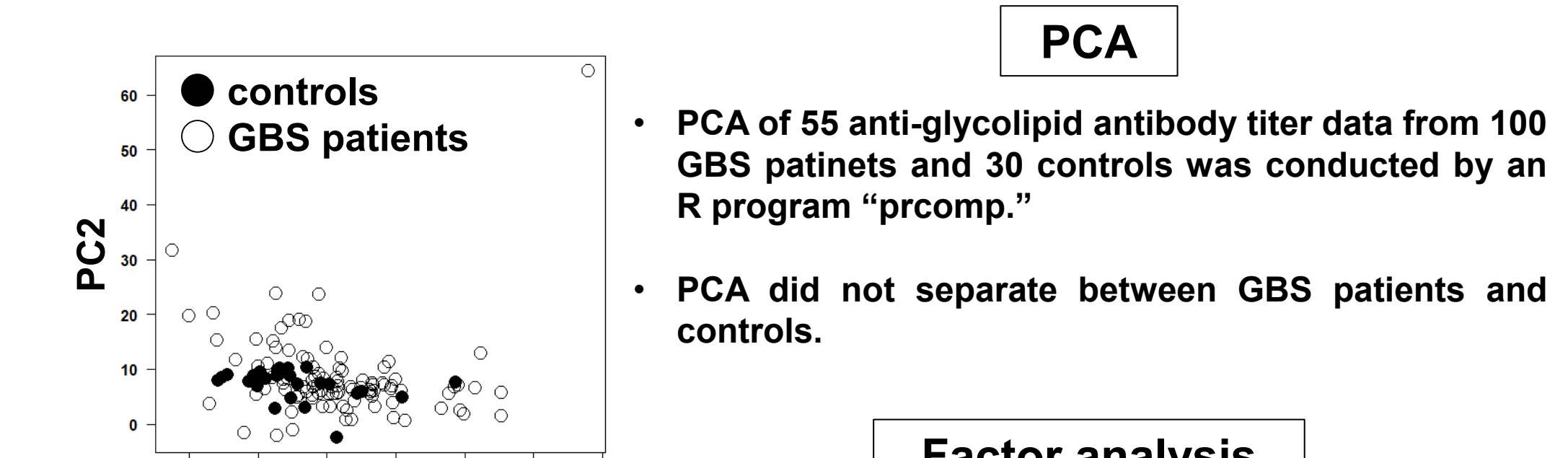
- PCA of gut bacteriome data from MS model clearly separated from control mice (Omura et al., 2020).
- PCA of gut virome data did not distinguish between the MS patients and healthy controls.

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.

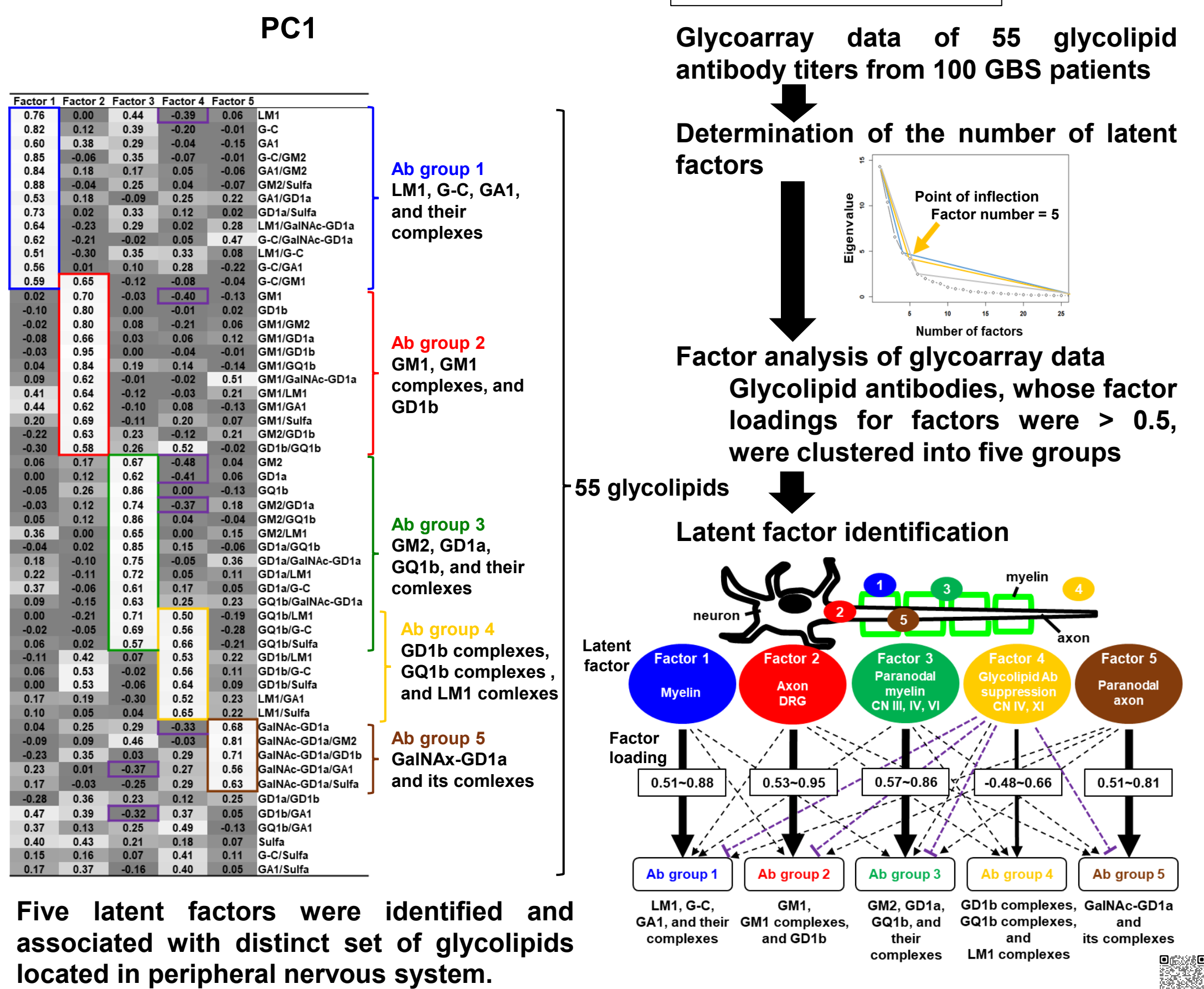


PCA vs. factor analysis of glycoarray data from the patients with Guillain-Barré syndrome



- PCA of 55 anti-glycolipid antibody titer data from 100 GBS patients and 30 controls was conducted by an R program "prcomp."
- PCA did not separate between GBS patients and controls.

Factor analysis



Five latent factors were identified and associated with distinct set of glycolipids located in peripheral nervous system.

Aim

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

Methods

The virome data were obtained as follows:

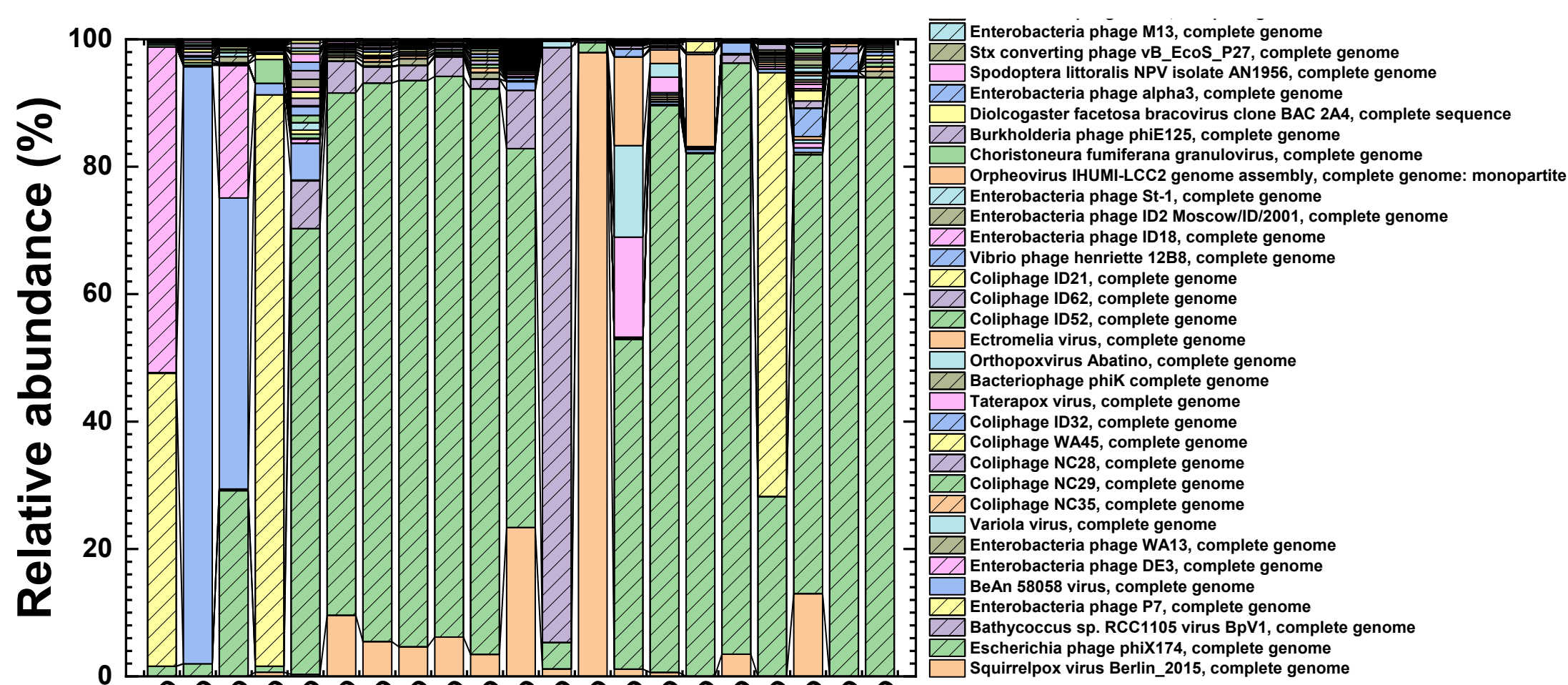
- Virome analysis of 22 lacrimal fluid samples using MinION device
- Virome data extracted from shotgun metagenome data (PRJNA391511, Fernandes et al., *J Pediatr Gastroenterol Nutr*, 2019) from 12 IBD patients and 12 healthy controls
- Virome data extracted from shotgun metagenome data (PRJNA1084182, 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

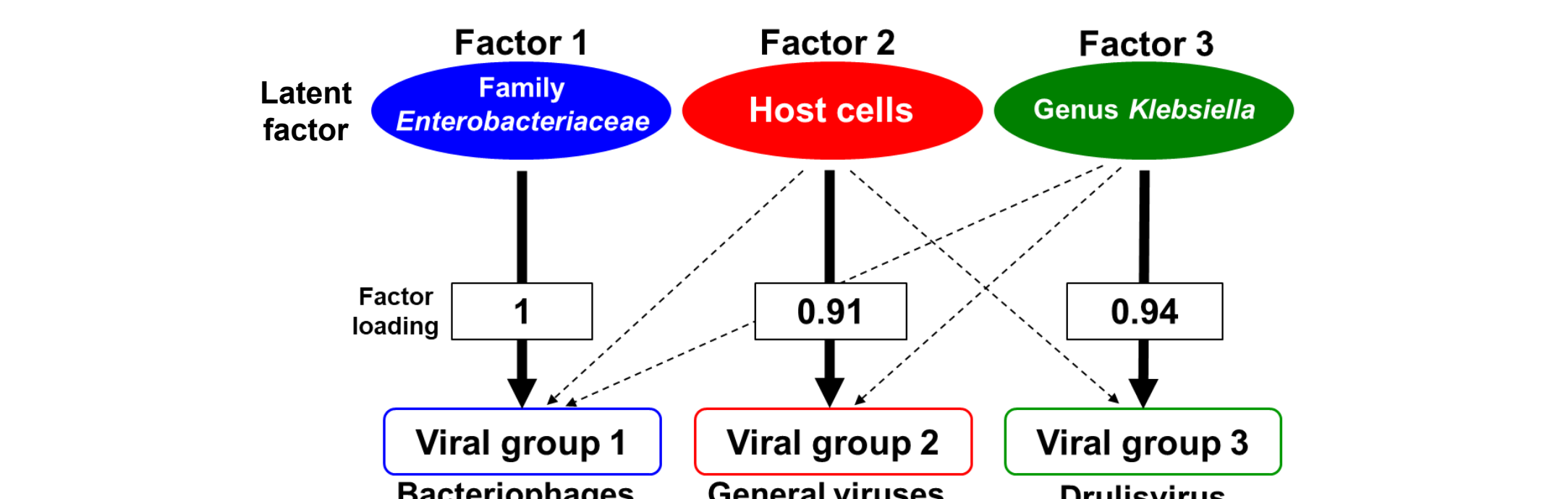
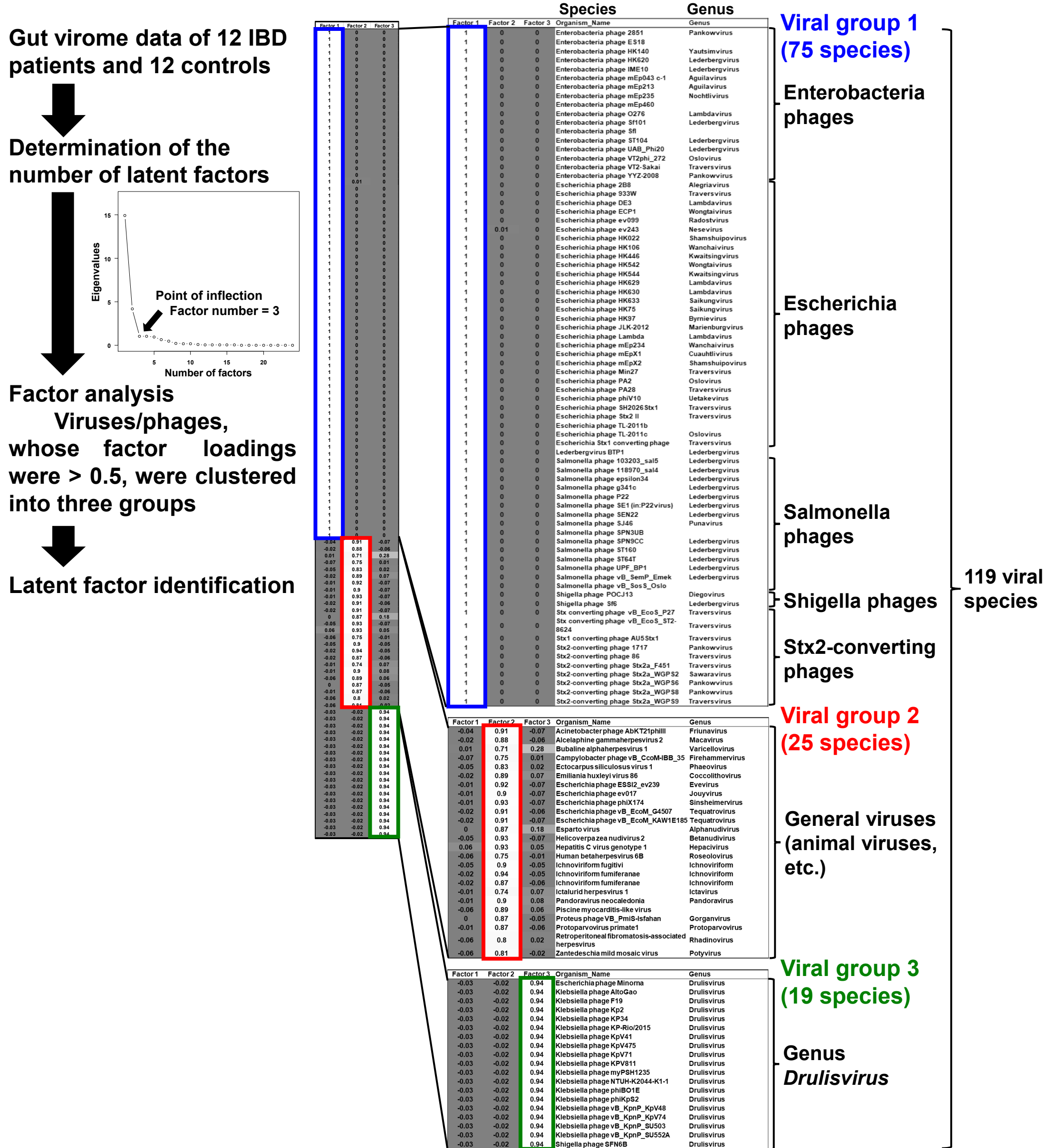
Virome is dominantly composed of bacteriophages (phages)



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

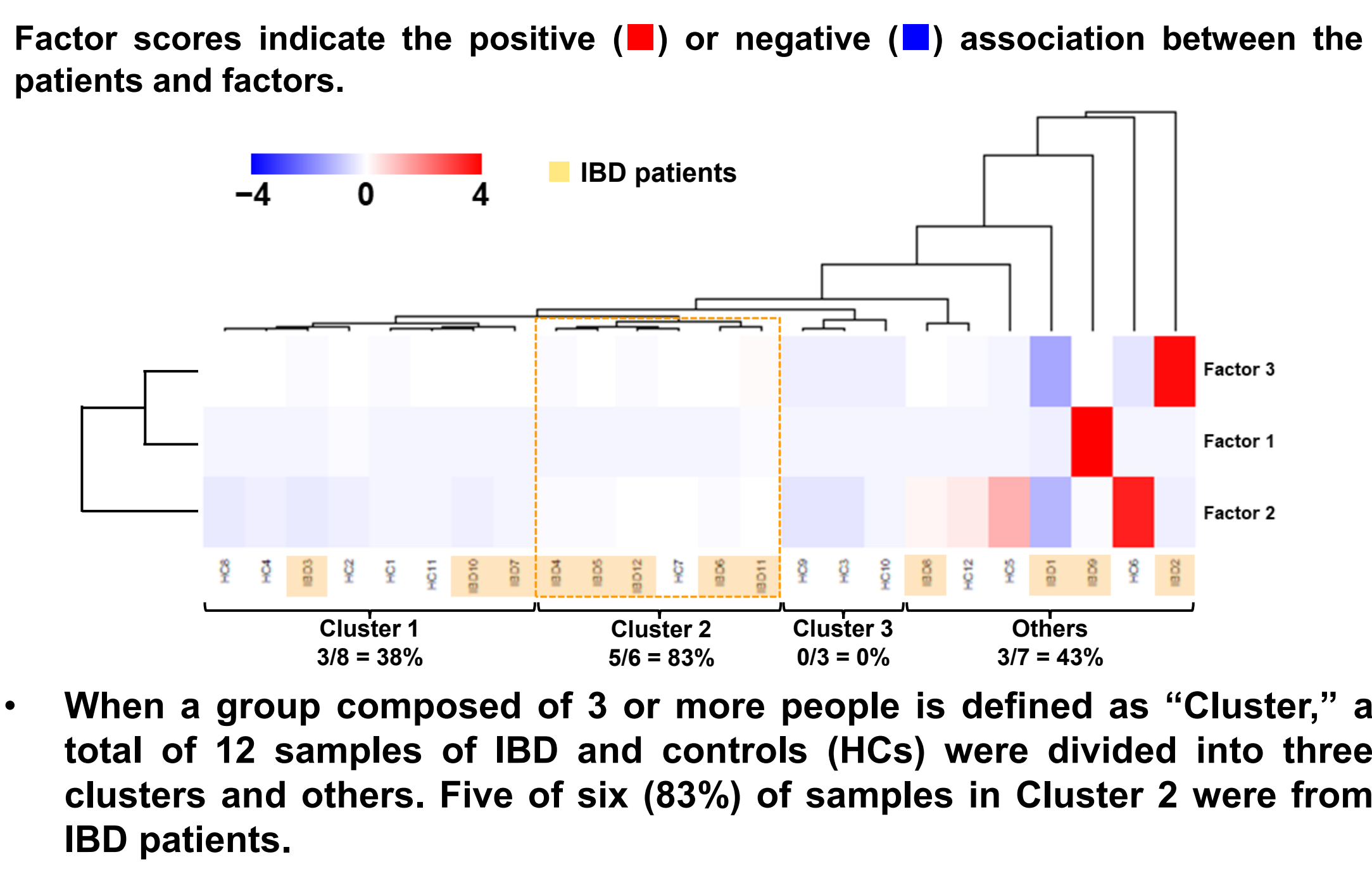
Factor analysis of gut virome in inflammatory bowel disease (IBD)

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



- 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



- 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.

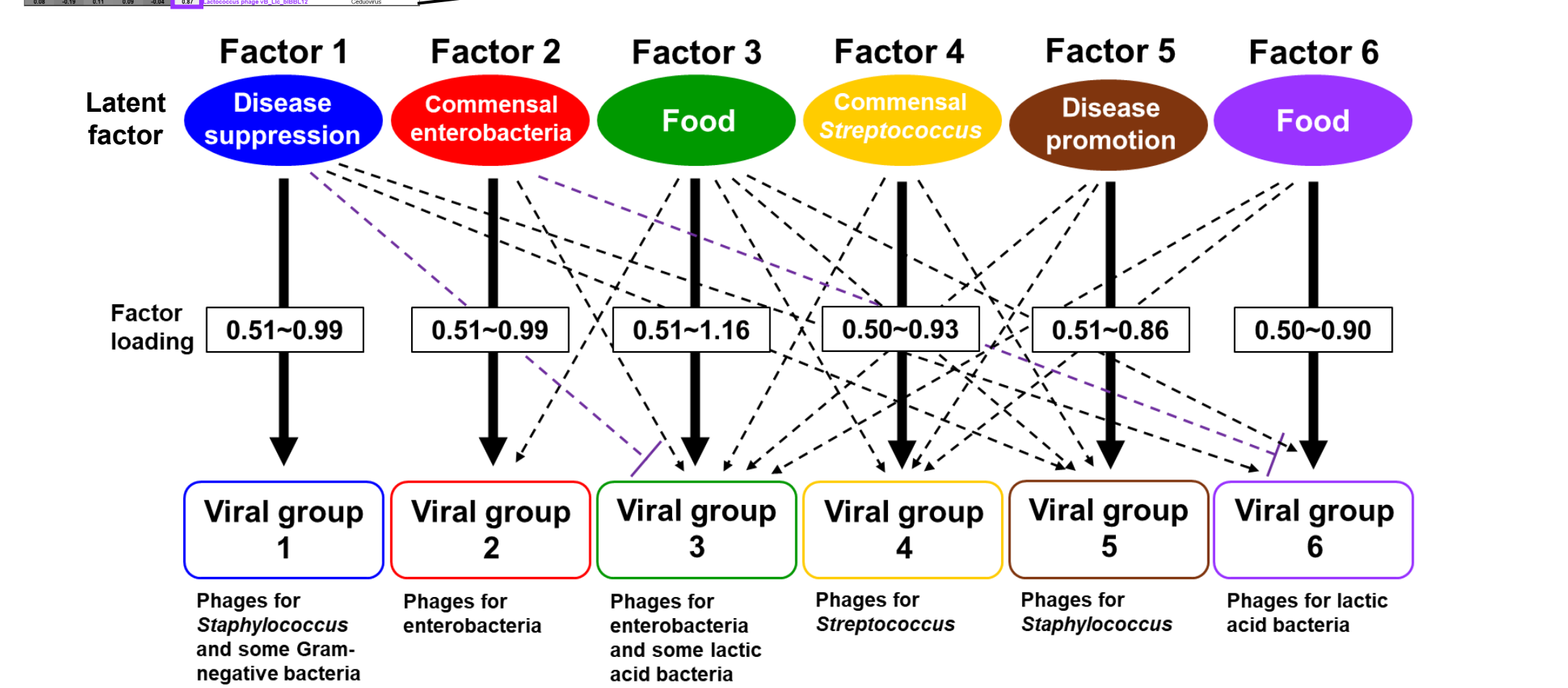
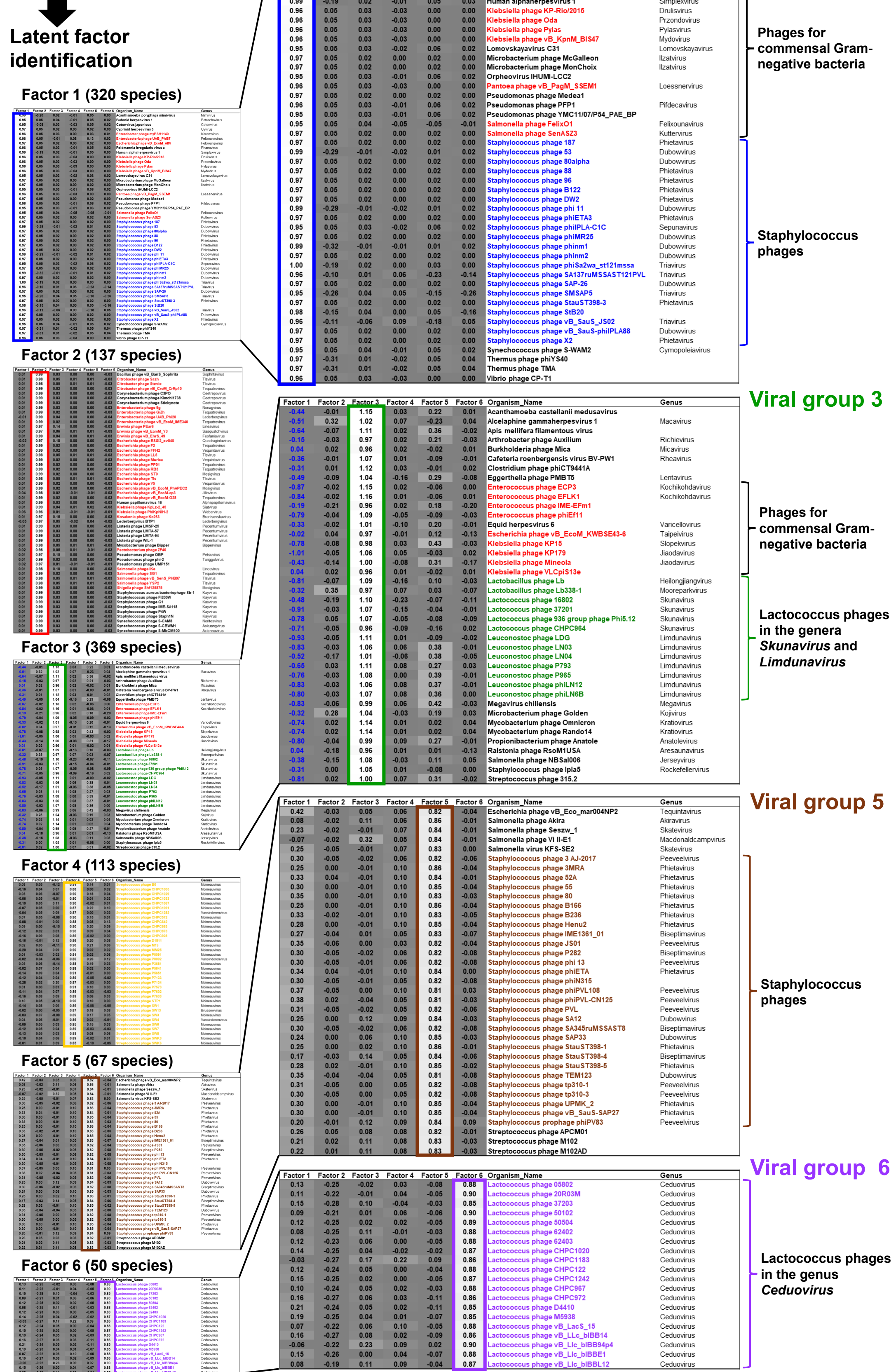
Criteria	RR-MS (n = 45)	HCs (n = 51)	P value
Age (mean ± SD)	41.8 ± 8.87	40.71 ± 14.69	P = 0.65
Age (<40 y: >40 y)	20 : 25	26 : 25	P = 0.54
Sex (Male:Female)	11 : 34	8 : 43	P = 0.31
Treatment (Yes:No)	35 : 10	0 : 51	

Determination of the number of latent factors

Factor analysis

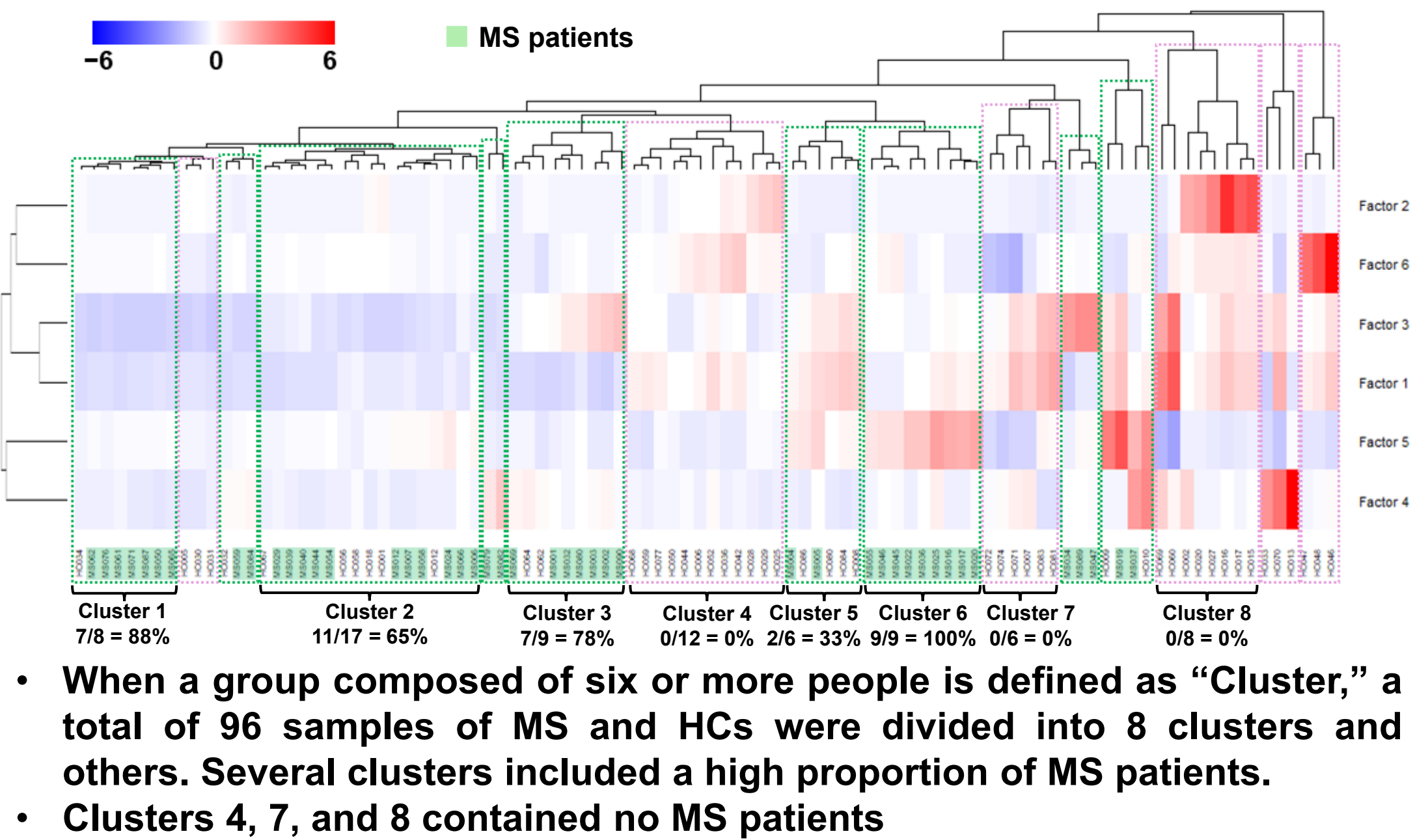
Viruses / phages, whose factor loadings were > 0.5, were clustered into six groups

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 did not work for the virome data.
- 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.