Machine learning model reveals hidden structure of human gut microbiome

Shion Hosoda
Suguru Nishijima
Tsukasa Fukunaga
Masahira Hattori
Michiaki Hamada

Video Byte

Keywords: Microbiome, metagenomics, latent Dirichlet allocation, LDA, human gut microbiome, enterotype, microbial assemblage, Bayesian model, machine learning, cluster analysis, co-occurrence, bioinformatics, partitioning around medoids, PAM, Dirichlet multinomial mixture, DMM, inflammatory bowel disease, IBD, butyrate, sequencing technology

Posted Date: November 3rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-102104/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Abstract

The human gut is home to a diverse community of microbes. Variations in the makeup of this community between individuals have been linked to diseases such as inflammatory bowel disease, diabetes, and cancer. Efforts to understand these differences have revealed three community types, or enterotypes, in humans, each representing the dominance of a single microbe. But because microbes co-mingle with many partners, studying the gut microbiome solely in terms of enterotypes misses on the highly nuanced nature of microbial interactions. Researchers recently addressed that shortcoming using a machine learning technique called latent Dirichlet allocation, or LDA. Their goal was to determine whether and how recurring microbial partnerships, or assemblages, are linked to the three enterotypes. Using gut metagenomic data gathered from 861 healthy adults across 12 countries LDA revealed three assemblages corresponding to each enterotype as well as a fourth wild-card assemblage that could be found in any gut. Understanding how these assemblages arise is crucial, as it could uncover properties of the gut microbiome that contribute to human disease.