[BIOINFO 2023]

HRGM v2: A global and high-quality human gut microbiome catalog of 155,211 near-complete reference genomes from 41 countries

Junyeong Ma[1], Nayeon Kim[1], Chan Yeong Kim[2], Sunmo Yang[1], Samuel Beck[3], and Insuk Lee[1,*]

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

[2] *European Molecular Biology Laboratory, Structural and Computational Biology Unit, 69117 Heidelberg, Germany*

[3] *Boston University School of Medicine, Boston, Massachusetts 02118, USA*

[*] *Corresponding author: insuklee@yonsei.ac.kr*

The human gut microbiome, often called the 'second human genome,' crucially impacts human health and diseases. Advances in assembly and binning methods using whole metagenome shotgun sequencing (WMS) reads have allowed us to access genome sequences from mostly uncultured gut microbiota. The Unified Human Gastrointestinal Genome (UHGG) catalog was the first comprehensive human gut microbiome catalog composed of these metagenome-assembled genomes (MAGs), but it was skewed towards MAGs from Western countries and China. To bridge this gap, we previously introduced the Human Reference Gut Microbiome (HRGM) catalog, emphasizing MAGs from underrepresented Asian countries. However, HRGM remained incomplete due to the lack of MAGs from South America, Asia, and Africa, and the presence of numerous species only consisting of very few medium-quality (MQ) genomes with potential false positives. Here, we present HRGM version 2, a global and high-quality human gut microbiome catalog comprising only near-complete (NC) genomes from 41 countries across continents. HRGM v2 contains 155,211 NC genomes and 4,824 NC species, which are 1.7 times and 1.5 times more than UHGG, respectively, accompanied by extensive functional and metabolic annotations. By integrating MAGs from non-Western countries, we shed light on the unique characteristics of their gut microbiomes. Moreover, HRGM v2 offers not only a Kraken 2 database but also a species-specific marker database and an optimized classification method for human gut microbiome analysis. HRGM v2 will be able to facilitate deeper insights into the human gut microbiome in future research.