[BIOINFO 2024]

The catalog of near-complete genomes improves taxonomic and functional understanding of the human gut microbiome

Junyeong Ma[1#], Nayeon Kim[1#], Jun Hyung Cha[1], Wonjong Kim[1],
Chan Yeong Kim[2], Sunmo Yang[1], Samuel Beck[3], and Insuk Lee[1,4*]

[1] *Department of Biotechnology, College of Life Science and Biotechnology, Yonsei University, Seoul 03722, Republic of Korea*
[2] *Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany.*
[3] *Department of Dermatology, Center for Aging Research, Chobanian & Avedisian School of Medicine, Boston University, Boston, USA.*
[4] *POSTECH Biotech Center, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea*
[#] *These authors contributed equally to this work*
[*] *Corresponding author: insuklee@yonsei.ac.kr*

A functional understanding of the human gut microbiome requires comprehensive genomic catalogs; however, existing catalogs lack sufficient geographic diversity and contain genomes of suboptimal quality. Most preexisting catalogs include a substantial number of medium-quality (MQ) genomes, missing up to 50% of genomic regions. These missing regions can significantly impact downstream functional analyses, potentially leading to incorrect hypotheses. To address these limitations, we developed the Human Reference Gut Microbiome (HRGM), a catalog of near-complete (NC) genomes (completeness ≥90% and contamination ≤5%) with expanded geographic sampling. HRGM includes 155,211 non-redundant NC genomes representing 4,824 prokaryotic species from 41 countries. Notably, through extensive sampling from Africa, we identified previously undetectable Africa-associated phyla, suggesting microbial links between humans and their environments. More importantly, the exclusive inclusion of NC genomes in HRGM provides a reliable landscape of functional components across taxonomic groups, enables accurate assessments of microbial metabolic capacities, and supports genome-wide metabolic modeling of entire microbiota to study metabolic interactions. These findings demonstrate that the exclusive cataloging of NC genomes provides novel functional insights into the human gut microbiome, which were not possible with previous catalogs containing largely incomplete genomes.