[BIOINFO 2025]

Exclusive cataloging of near-complete genomes unveils novel functional insights into the human gut microbiome

Junyeong Ma[1#], Nayeon Kim[1#], Jun Hyung Cha[1], Wonjong Kim[1], Chan Yeong Kim[2], Yong-ho Lee[3], Han Sang Kim[4], Yoon Dae Han[5], Dongeun Yong[6], Eugene Han[7], Sunmo Yang[1], Samuel Beck[8], and Insuk Lee[1,9*]

[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 Internal Medicine, Institute for Endocrine Research, Institute for Innovation in Digital Healthcare (IIDH), Yonsei University College of Medicine, Seoul 03722, Republic of Korea
[4]Yonsei Cancer Center, Division of Medical Oncology, Department of Internal Medicine, Graduate School of Medicine, Brain Korea 21 Project, Yonsei University College of Medicine, Seoul 03722, Republic of Korea.
[5]Division of Colorectal Surgery, Department of Surgery, Yonsei University College of Medicine, Seoul 03722, Republic of Korea
[6]Department of Laboratory Medicine, Research Institute of Bacterial Resistance, Yonsei University College of Medicine, Seoul 03722, Republic of Korea
[7]Division of Endocrinology, Internal Medicine, Keimyung University School of Medicine, Daegu 42601, Republic of Korea
[8]Department of Dermatology, Center for Aging Research, Chobanian & Avedisian School of Medicine, Boston University, Boston, MA 02118, USA.
[9]DECODE BIOME Co., Ltd. Incheon 21983, Republic of Korea
[#]These authors contributed equally to this work

*Corresponding author:
Insuk Lee
Department of Biotechnology, College of Life Science and Biotechnology, Yonsei University, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea
Tel: +82-10-4186-8706, E-mail: insuklee@yonsei.ac.kr

Abstract

Understanding the human gut microbiome requires comprehensive genomic catalogs, yet many lack geographic diversity and contain medium-quality metagenome-assembled genomes (MAGs) missing up to 50% of genomic regions, potentially distorting functional insights. To overcome this, we developed an enhanced Human Reference Gut Microbiome (HRGM2), a catalog of near-complete (NC) MAGs (≥90% completeness, ≤5% contamination, GUNC clade separation score <0.45) and isolate genomes (collectively "NC genomes") with extensive geographic coverage. HRGM2 comprises 155,211 non-redundant NC genomes from 4,824 prokaryotic species across 41 countries, representing a 66% increase in genome count and a 50% boost in species diversity over the NC genome-based Unified Human Gastrointestinal Genome catalog. Given the high taxonomic complexity of the gut microbiota, DNA-based species profiling using HRGM2 outperforms marker-based methods in accuracy. Notably, with exclusive use of NC genomes, HRGM2 improves metabolic capacity assessment and enables high-confidence, automated genome-scale metabolic models (GEMs) of entire microbiota. These findings suggest that NC genome catalogs improve reliability of microbiome functional insights.