

Human reference gut microbiome catalog including newly assembled genomes from under-represented Asian metagenomes



Chan Yeong Kim^{1†}, Muyoung Lee^{1†}, Sunmo Yang¹, Kyungnam Kim², Dongeun Yong², Hye Ryun Kim³, and Insuk Lee^{1*}

¹Department of Biotechnology, College of Life Science and Biotechnology, Yonsei University, Seoul 03722, Korea. ²Department of Laboratory Medicine, Research Institute of Bacterial Resistance, College of Medicine, Yonsei University, Seoul 03722, Korea ³Division of Medical Oncology, Department of Internal Medicine, Yonsei Cancer Center, College of Medicine, Yonsei University, Seoul 03722, Korea

BACKGROUND

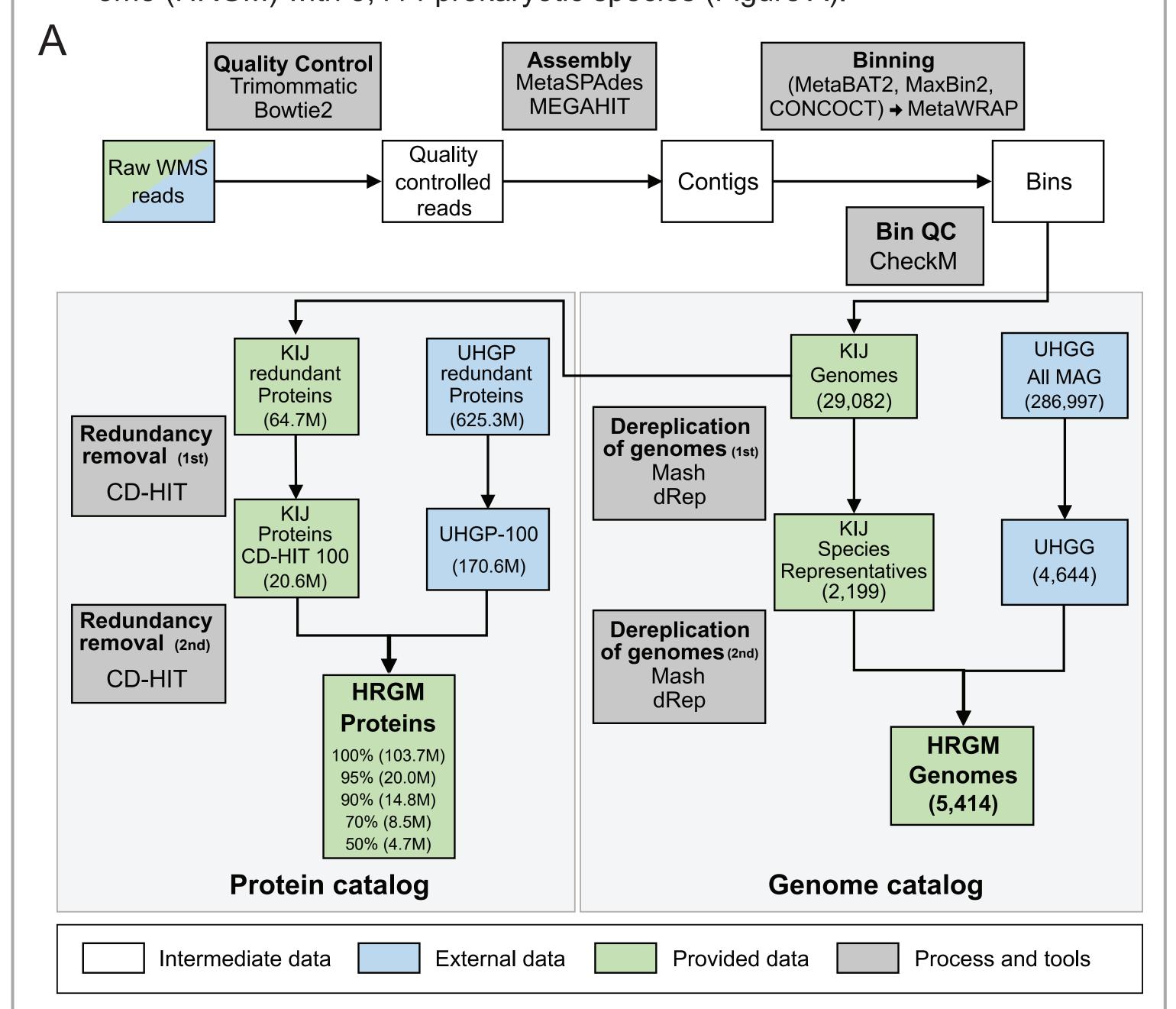
■ A large portion of human gut microbiota are un-isolated yet; Therefore, we lack genome sequences of them.

Underlined: Presenting author, †These authors contributed equally to this work, *Corresponding author: insuklee@yonsei.ac.kr

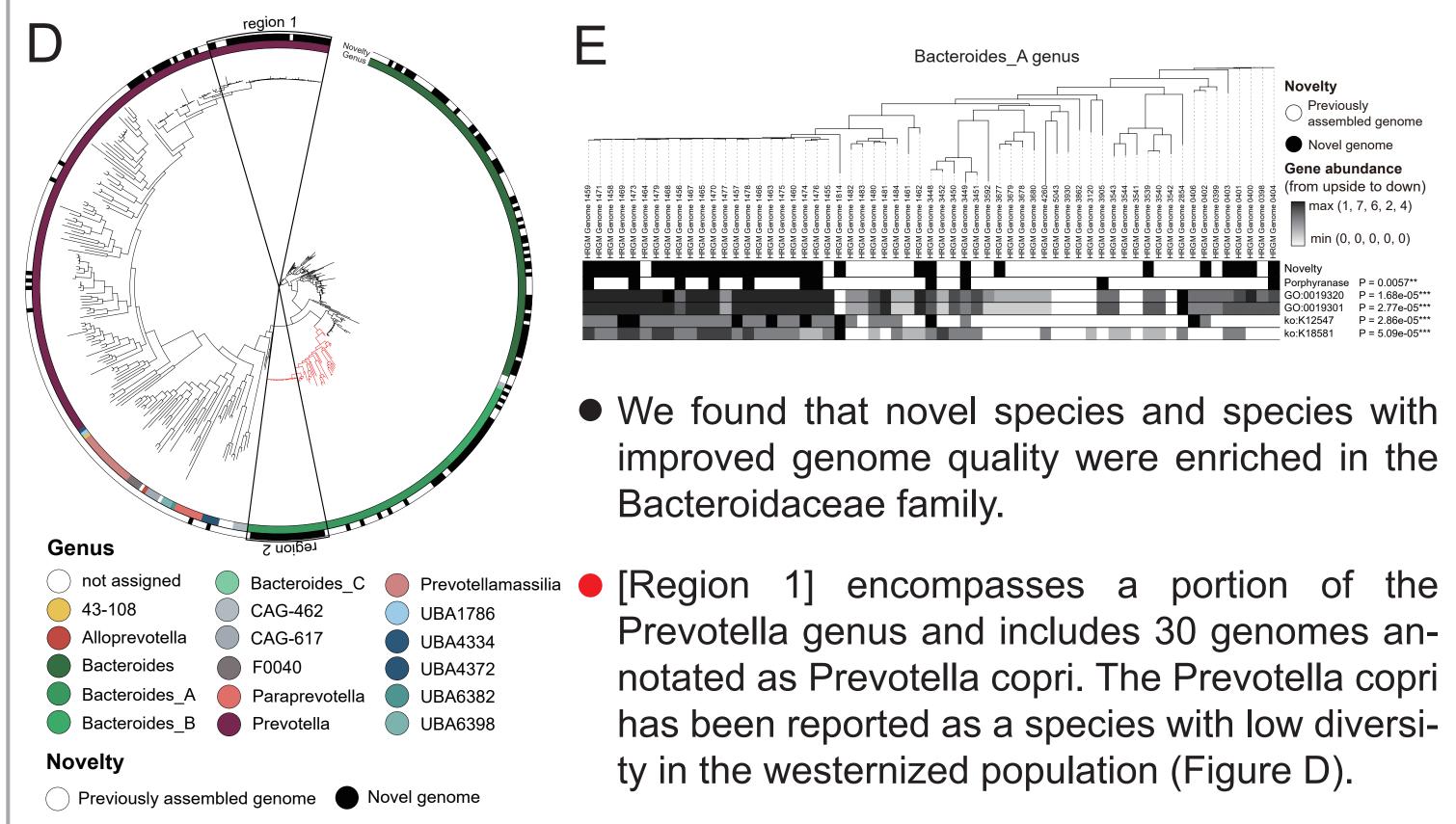
- Recently, metagenome-assembled genomes (MAGs) substantially increase the phylogenetic diversity of uncultured taxa; However, it is biased for geographical location and lifestyle.
- In this study, we constructed comprehensive human gut microbiome genome and protein catalogs by incorporating fecal whole metagenomic shotgun (WMS) sequencing samples from under-represented populations.

METHODS

- We compiled fecal WMS sequencing of 90 Koreans (newly sequenced in this study), 805 Japanese, and 110 Indian (Public dataset) (Figure A).
- For the 90 Korean samples, we performed ultra-deep (25Gbp 60Gbp, average 30Gbp) WMS sequencing.
- We performed de-novo metagenomic assembly and species-level dereplication with UHGG genomes, and constructed Human Reference Gut Microbiome (HRGM) with 5,414 prokaryotic species (Figure A).



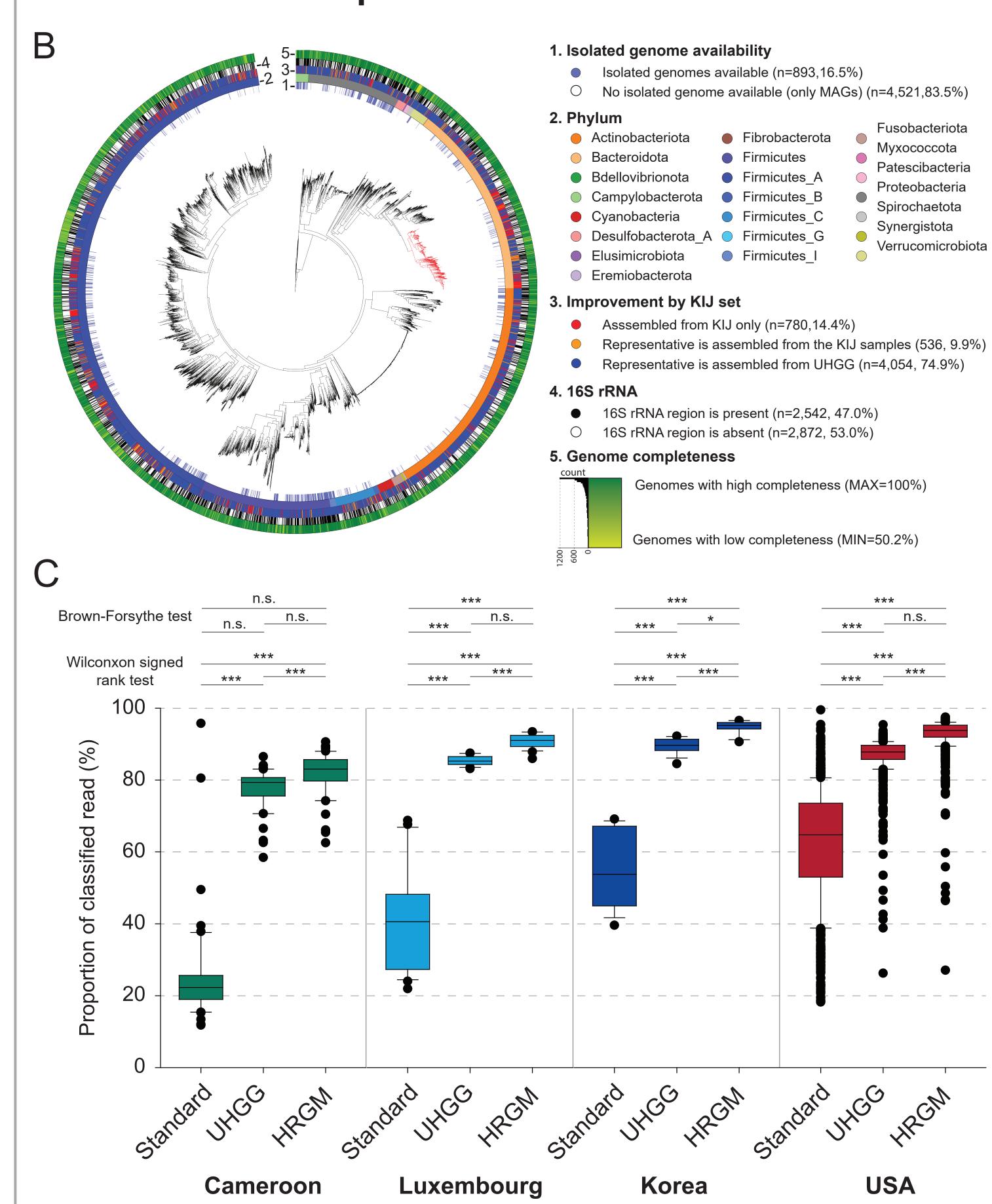
RESULT 2 - Novel Bacteroidaceae are associated to diet-related lifestyle of Asian



- [Region 2] is in Bacteroides_A genus, and 22 genomes are annotated as Bacteroides_A plebeius (Figure D).
- Japanese B. plebeius have been reported to harbor a Porphyranase gene that transferred from marine microbe and contributed to the metabolism of agal carbohydrates.
- We found Porphyranase and other carbohydrates metabolism-related functional categories are enriched in the newly assembled Bacteroides A (Figure E).

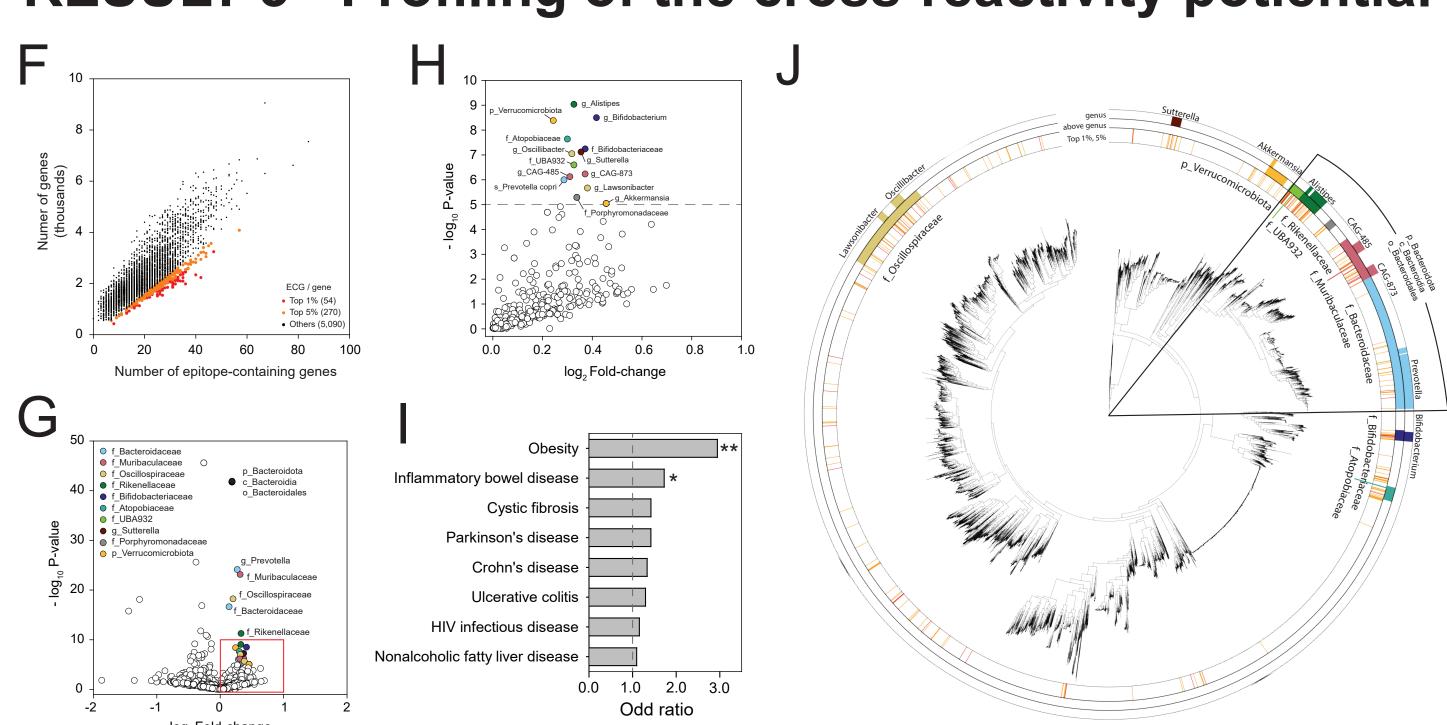
RESULT 1 - The improvement of HRGM

HRGM is available at https://www.mbiomenet.org/HRGM/



- We obtained 780 novel species and improved the genome quality of 536 species representatives (Figure B).
- We found that 4,521 (83.5%) species are exclusively assembled from metagenomic data (Figure B).
- We observed significant improvement in classification proportion compared to the RefSeq-CG-based standard database and the UHGG-based database, regardless of ethnicity (Figure C).
- Moreover, variance in classification rate across samples is reduced except for the Cameroon cohort (Figure C).

RESULT 3 - Profiling of the cross-reactivity potiential



- We screened human autoimmune disease (AID) related epitope sequences from the entire human gut microbiome level with HRGM, and identified genomes with many epitope-containing genes (ECGs) (Figure F).
- We found genomes with many ECGs are not randomly dispersed in the tree but enriched in specific clades (Figure J).
- We identified taxa with higher epitope sequences (Figure G, H). Those taxa include *Akkermansia*, *Prevotella*, Rikenellaceae, which previously reported as human AID-related taxa.
- Obesity and IBD were significantly more associated with high cross-reactivity taxa, which suggest their potential role in inflammation.