HumGut: A comprehensive database of prokaryotic genomes in the healthy human gut

Pranvera Hiseni
Knut Rudi
Robert C. Wilson
Finn Terje Hegge
Lars Snipen

Video Byte

Keywords: human gut microbiome, genome collection, database, prokaryote, microbiota, Microbiome, gut microbe, metagenome, genome, metagenomics, genomics, reference database, sequence alignment, taxonomic classification, taxonomy, phylogenetics, nomenclature, species identification, bacteria, read mapping, sequence annotation

Posted Date: October 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-967259/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Genomics research has greatly increased understanding of the human gut microbiome, but the existing reference databases remain insufficient, failing to map up to half of the sequences obtained in human gut studies. To solve this problem, researchers recently created HumGut, a comprehensive global reference database for the genomes of gut microbes in healthy humans. The researchers built the database by comparing nearly half a million publicly available prokaryote genomes with over 3,500 gut metagenomes from healthy humans worldwide, and retaining the prokaryote genomes that closely matched the sequences in healthy human guts. HumGut was approximately the same size as the recently released UHGG collection and half the size of a standard reference database. However, HumGut outperformed both other databases in classifying metagenomic reads from human gut samples, resulting in a lower percentage of unclassified reads. HumGut also effectively mapped metagenomes from patients with inflammatory bowel disease, indicating that it is useful even for microbiomes that differ from those used to build the database. Although the species naming system in HumGut needs to be refined, this new database will help researchers identify the majority of human gut microbiome members in any study they conduct worldwide.