Faculty of Dentistry, Oral & Craniofacial Sciences Blog: the Centre for Host-Microbiome Interactions

Human Microbiome Atlases

The key to advancing our understanding of the critical role played by the microbiome in health and disease is access to data from a wide range of studies and cohorts. However, still there is a lack of integrative functional and compositional analysis across the cohorts and regions, together with rigorous application of machine learning to provide a mechanistic understanding of the microbiome and biomarker identification.

To provide a central public resource for exploring the microbiome in different settings, we performed a large-scale integrative analysis of 6,014 publicly available shotgun metagenomics stool samples. Samples with at least ten million high-quality sequencing reads from healthy and disease cohorts from nineteen different countries across five continents were selected. We performed pan-metagenomic analysis and created the Human Gut Microbiome Atlas (HGMA) (**www.microbiomeatlas.org**) using quantitative analysis of shotgun metagenomics based on microbial genomes assembled using "Metagenomic Species Pan-genomes (MSPs)". We also extensively characterized the functions and phenotype of the identified MSPs in seven different categories: KEGG orthologs, protein families, carbohydrate-active enzyme, antimicrobial resistance, microbial phenotype, virulence factor, and biosynthetic gene clusters, as shown in the accompanying figures.

Using our HGMA resource and our integrative platform, cluster analysis revealed functions commonly enriched in disease that provide their bacterial carriers increased fitness to thrive in altered conditions. Such functional adaptations play indirect roles in disease pathology for example, allowing bacteria to use additional carbon sources or increase their ability to survive environmental stresses.

In addition to functional bacterial signatures, we established machine learning models for extracting disease signatures of the gut microbiome, based on SHAP (SHapley Additive exPlanations) values, leading to the identification of microbial markers of diseases, which produced a novel species-level biomarker for pan-cancer and other diseases in our atlas. These models were able to classify between a pool of randomly selected healthy samples and disease groups with variable discriminatory performances. The computer scripts for the functional clusters, SHAP calculation, plotting, enrichment in disease/region are publicly available at https://github.com/sysbiomelab.

Following the creation of our HGMA resource, we are developing a new Human Oral Microbiome Atlas (HOMA), which consists of more than five thousand publicly available and newly generated samples. HOMA will cover more than ten different

diseases across several countries with features for bacterial functional analysis, including antimicrobial resistance.

Large-scale integration of shotgun metagenomics association studies provides novel information on functional changes in the microbiome and host physiology and even new microbiome-based biomarkers, treatments, and therapies.

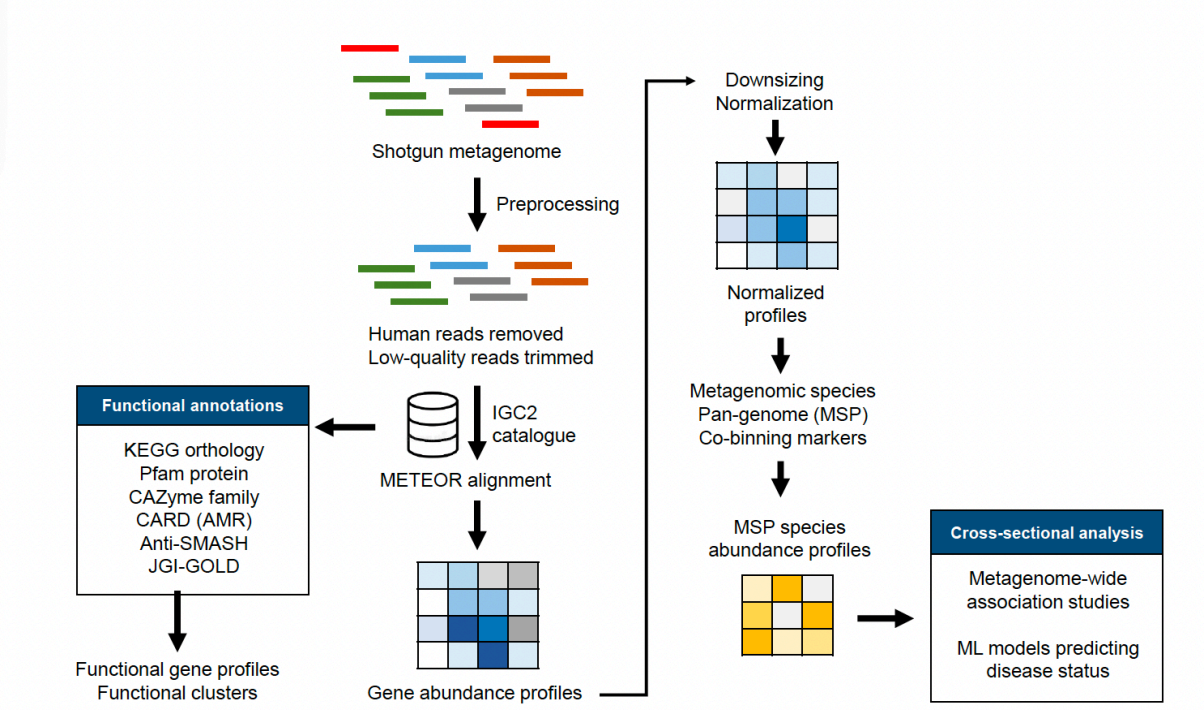

Figure 1. Workflow for development of the microbiome atlas.

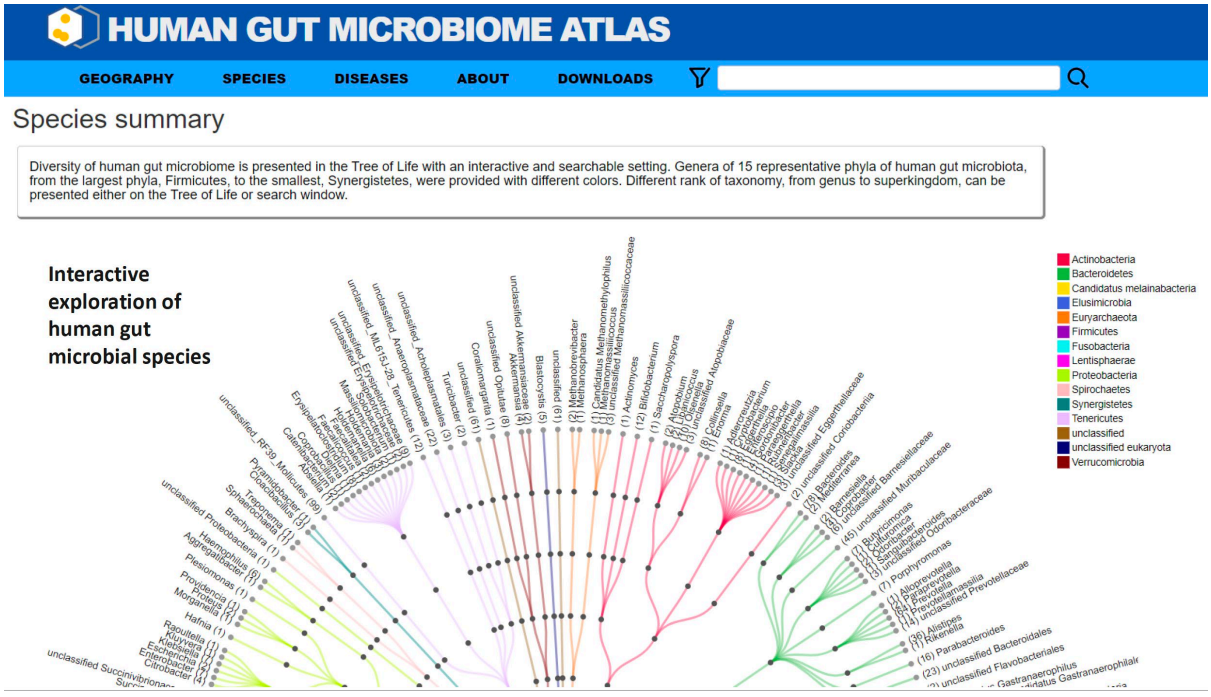

Figure 2. A snapshot of the microbiome atlas page for exploring the species using the search engine or tree of life.