

Statistical advances help unlock mysteries of the human microbiome

August 13 2015



People from different continents harbor distinct bacterial strains. Gut microbiomes from five continents (red flags) contain strains with significantly different genes, as shown by their phylogenetic relationships (tree superimposed on the map with black lines indicating relatedness of the bacteria). Europe (green) and North America (blue) share two groups of strains, which are distinct from those in China (yellow), South America (purple) and Africa (red). Credit: Katherine Pollard, Gladstone Institutes

Advances in the field of statistics are helping to unlock the mysteries of

the human microbiome—the vast collection of microorganisms living in and on the bodies of humans, said Katherine Pollard, a statistician and biome expert, during a session today at the 2015 Joint Statistical Meetings (JSM 2015) in Seattle.

Pollard, senior investigator at the Gladstone Institutes and professor of epidemiology and biostatistics at the University of California, San Francisco, delivered a presentation titled "[Estimating Taxonomic and Functional Diversity in Shotgun Metagenomes](#)" during an invited session focused on statistics, the [microbiome and human health](#).

"While we are only just beginning to understand the complex roles microbes play in human biology, it is clear specific changes in microbial flora are associated with—and sometimes cause or cure—disease in the host," said Pollard while explaining her research focus. "Some of the best-supported links are with autoimmune diseases, which are on the rise in the United States, perhaps due to antibiotic use and lack of exposure to a diverse collection of microbes during childhood. This 'hygiene hypothesis' suggests that health risks not attributable to human genetics and behavior may stem from differences in microbiome composition between individuals."

What's more, a given [microbial species](#) can share less than 50% of the same genes when found in two people. These differences track with functional capabilities of the microbial communities, including genes related to sugar metabolism, biosynthesis and two-component systems. "Two people with the exact same species of bacteria in their guts could experience very different interactions with these bacteria because different strains simply are not doing the same thing," said Pollard.

For this reason, it is important to determine not only the types of microbes present in a given sample, but also the genetic makeup of each strain. However, this presents a considerable Big Data challenge,

requiring advances in statistical methodology and new software for accurate analysis of metagenomics data. "The development of metagenomic sequencing of the total DNA in a microbial sample from the human body has allowed us to estimate the abundance of specific microbes and microbial genes. But, as with any new technology, metagenomics has many biases and errors that must be corrected analytically before we can accurately compare data across samples," said Pollard. "This has limited our understanding of both the extent and impact of microbial variation in many environments, most importantly the human microbiome."

Metagenomics poses many analysis challenges, from errors reading DNA sequences to decoding which sequences come from which of the hundreds of microbial species in a microbiome sample. One of the biggest issues is many of the microbial strains in a given person have never been sequenced. Even in the well-studied human gut microbiome, it was estimated, on average, 43% of species abundance could not be captured by available microbial reference analysis methods.

To address this and other microbiome research problems, Pollard and Stephen Nayfach, a bioinformatics graduate student in Pollard's group at the Gladstone Institutes, developed a suite of new statistical software to rapidly and accurately estimate the presence and function of microbes in a metagenome. Their programs—called MicrobeCensus, ShotMAP and PhyloCNV—made significant methodological improvements that allowed the scientists to accurately quantify the specific strains in the [human microbiome](#) using sequencing reads as short as 50 base pairs.

Using the new tools, Pollard's lab investigated a reported finding that obese people have a lower ratio of bacteria from the phylum Bacteroidetes to bacteria from the phylum Firmicutes compared with lean individuals. Although the scientific literature and the general media had heralded this association as noteworthy, several reports questioned

its existence.

To test the validity of the association, Pollard's group conducted an extensive assessment of the relationship between body mass index (BMI) and the taxonomic composition of the gut microbiome. Their meta-analysis of data from multiple studies did not find a significant association between BMI and the relative abundance of any bacterial species, said Pollard during her presentation.

She said new statistical advances will enable scientists to perform other forms of microbiome research such as identifying microbial species and genes that are biomarkers for disease onset or conducting drug development that targets the microbiome.

"The microbiome clearly plays a role in host biology, but this role is complex and must be analyzed in the context of diet, drugs and host genetics," she concluded.

Provided by American Statistical Association

Citation: Statistical advances help unlock mysteries of the human microbiome (2015, August 13) retrieved 3 May 2024 from

<https://phys.org/news/2015-08-statistical-advances-mysteries-human-microbiome.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.