MICROBIOME

Microbiome genome structure drives function

Differences in microbial genomes can result in vastly different phenotypes and functions. Consequently, it is critical to understand the genome variations that differentiate microbial strains. Here, we discuss recent exciting advances that enable structural variant measurement, their associated phenotypes and the horizon for future discovery.

Matthew G. Durrant and Ami S. Bhatt

The microbiome field has quickly progressed from 16S ribosomal RNA gene sequencing, providing insight into the taxonomic composition of a community based on a single gene, to deep metagenomic sequencing — the simultaneous sequencing of multiple microbial genomes in a complex community, which generates a wealth of genetic data describing complex communities. Taking full advantage of these data presents technical challenges, but the main hurdle to making the most of these rich datasets can be our own creativity and determination to maximize what we can learn from them.

In a recent issue of *Nature*, Zeevi et al. seized an opportunity to use short-read metagenomic sequencing to explore a poorly understood aspect of these datasets — structural variation[1]. Structural variation in microbial genomes is generally more difficult to detect than other genomic variations, such as small single nucleotide polymorphisms and small insertions or deletions ('indels'). This is true of sequenced isolates, and even more so for metagenomes. As sequencing technologies improve, offering us longer reads and better resolution to detect large structural variation, it is tempting to wait for these developments before diving into the world of structural variation in microbial genomes. However, Zeevi and colleagues took full advantage of short-read sequencing data, and their findings shed new light on how important structural variations can be, especially in the context of human health (Fig. 1).

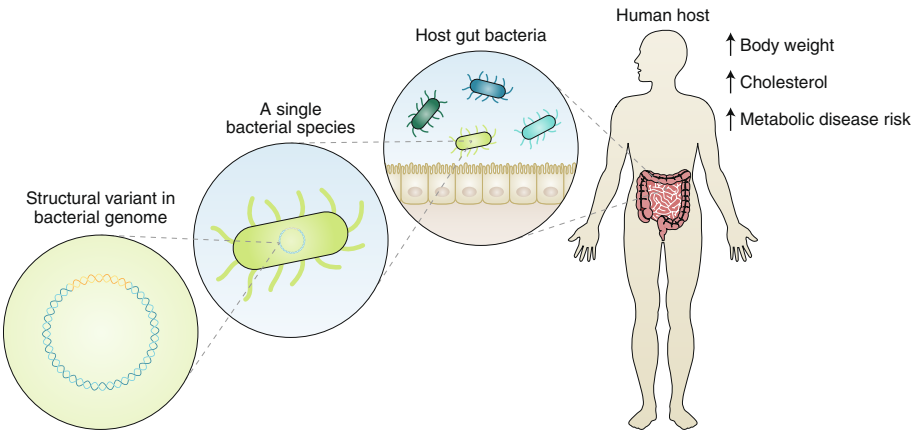

Fig. 1 | Structural variation in the gut microbiome. Structural variations at the level of individual microbial genomes in the human gut microbiome associate with metabolic phenotypes in the human host, including body weight, cholesterol and other markers of metabolic disease risk.

Zeevi and colleagues first tackled an important problem — assigning sequenced reads to the genome from which they originate. Typically, reference databases are limited to genomes that meet a certain coverage cut-off, which requires a reference genome collection to be built for each sample and still leaves one questioning how confidently a given read has been assigned to a particular genome. In this study, the authors instead developed an iterative coverage-based read assignment algorithm that repeatedly adjusts read assignments using relative species abundance within the sample. These corrected read alignments were used to thoroughly characterize structural variation in the microbiome. Another algorithm, named SGV-Finder, was developed to identify structural variants (SVs) that are either completely absent in many samples (deletion SVs) or those whose coverage is highly variable across samples (variable SVs). Overall, they identified 2,423 variable SVs and 5,056 deletion SVs in 56 common and abundant microbiome species. They detected at least one, and often many, structural variations in every subject they analysed, indicating the great variety of SVs across individuals.

This tool was also applied to independent samples from the Dutch Lifelines DEEP cohort, where over 70% of identified SVs were replicated. These SVs were highly stable within individual microbiomes over long time periods, and cohabiting individuals were much more likely to share such SVs relative to individuals who do not live together. This suggests that SVs may be transmitted between individuals, much in the same way that microbial taxa are.

SV regions were predicted to encode 'accessory' genes and operons that can be laterally transferred between species, as opposed to core genes that are vertically inherited during the evolution of a given taxon. In keeping with this expectation, the authors found that the SV regions are depleted of housekeeping modules and strongly enriched for transporters such as ABC-2 type systems and type-IV secretion system modules, suggesting that SVs play an important role in bacterial adaptation. Indeed, the gene content of these SVs suggests that they are part of the vast 'mobilome' (that is, the accessory genome) of microorganisms.

Perhaps unsurprisingly, SVs were also enriched for genes with no assigned function. Using more sensitive annotation approaches, the authors found that many of these were associated with bacteriophage, plasmids, transposons, antibiotic-producing genes and CRISPR-associated genes. This indicates that structural variation is likely driven by horizontal gene transfer, carrying valuable molecular tools that are spreading through the microbial communities in our guts. Using a method to estimate bacterial

growth rates from single metagenomic samples[2], 44 deletion SVs were strongly associated with bacterial growth rates, suggesting that these SVs may affect bacterial fitness.

Next, the authors compared SVs with important biomarkers of health, such as blood pressure, cholesterol levels, waist circumference, body mass index and glucose levels, resulting in over 100 significant associations. Impressively, one-third of these associations were replicated in an independent cohort. Strong associations between SVs and traits, such as high-density lipoprotein cholesterol and body weight, were identified with one SV in *Anaerostipes hadrus* coding for a composite inositol catabolism-butyrate biosynthesis pathway that was associated with lower risk of metabolic disease. That is, it is not simply the species of bacteria within your gut that may affect your weight, cholesterol and risk of disease, but also the specific SVs that are harboured in those species' genomes.

But why focus on SVs? In addition to species composition, knowledge of SVs holds promise in helping our field answer perplexing questions related to the microbiome. For example, we know that the gut microbiome influences the efficacy of programmed cell death 1-based immunotherapies[3–5], but different groups have found that different taxa are associated with these responses. Including an analysis of SVs may help to clarify the mechanisms underlying these interesting and clinically promising associations. These SVs highlight the flexibility of microbial genomes in a metagenomic context and they encode for genes that are involved in clinically relevant phenotypes, such as antibiotic resistance, virulence, and adaptation to different nutrients, growth conditions and environments.

Recently, new read-cloud and long-read sequencing approaches have become available[6,7], which could make detecting and characterizing these large structural variations easier. Tools such as these will help us to better determine the genomic context of SVs through more high quality genome assemblies. Interestingly, while many SVs encode large, horizontally transferred genomic fragments consisting of multiple genes, such as transposases and integrases, there are SVs that may be as small as tens of base pairs. In the case of these smaller elements, their gene content may not be as informative as their site of insertion, which can act as an important source of mutational variation[8].

As metagenomic methods continue to develop, these datasets will only increase in their value to researchers and clinicians alike. The human gut is a vast reservoir of genetic potential, and we are learning more and more about how it relates to human health and, in turn, how it may be manipulated in a therapeutic context. Many questions remain going forward, such as the exact mechanism of each SV–bacteria–host interaction, but these recent findings provide us with a valuable resource and have certainly inspired the field by opening researchers' minds to a new avenue of scientific inquiry. ❐

Matthew G. Durrant[1] and Ami S. Bhatt[1,2]*

[1]*Department of Genetics, Stanford University, Stanford, CA, USA.* [2]*Division of Hematology, Blood and Marrow Transplantation, Department of Medicine, Stanford University, Stanford, CA, USA.*
**e-mail: asbhatt@stanford.edu*

Published online: 22 May 2019
https://doi.org/10.1038/s41564-019-0473-y

References

1. Zeevi, D. et al. *Nature* **568**, 43–48 (2019).
2. Korem, T. et al. *Science* **349**, 1101–1106 (2015).
3. Routy, B. et al. *Science* **359**, 91–97 (2018).
4. Gopalakrishnan, V. et al. *Science* **359**, 97–103 (2018).
5. Matson, V. et al. *Science* **359**, 104–108 (2018).
6. Bishara, A. et al. *Nat. Biotechnol.* **36**, 1067–1075 (2018).
7. Moss, E. L. & Bhatt, A. S. *bioRxiv* https://doi.org/10.1101/489641 (2018).
8. Durrant, M., Li, M. M., Siranosian, B. & Bhatt, A. S. *bioRxiv* https://doi.org/10.1101/527788 (2019).

Competing interests

The authors declare no competing interests.