

Sequence-based identification of human-associated microbiota

Credit: Sean Prior / Alamy Stock Photo



Over the past twenty years, advances in DNA sequencing technology have deepened our knowledge in every niche of biology, from annotation of the human genome to sequencing the human-associated microbial metagenome, the genetic material of the microorganisms that inhabit almost every surface of the human body.

Study of the human microbiota historically relied on culture-dependent methods to isolate and grow bacterial colonies in a predetermined medium. The inability to cultivate a large portion of microorganisms, however, substantially underestimated the biodiversity of human-associated microbial communities.

An early milestone in this field was the adoption of a technology, previously pioneered by Carl Woese, Norman Pace and others to identify environmental bacteria, that was based on sequencing small subunit ribosomal RNA genes (16S rRNA). Using this approach, Wilson and Blitchington compared the diversity of cultivated and noncultivated bacteria within a human faecal sample in 1996. Since then, sequencing of 16S rRNA genes from complex communities has become a powerful tool for assessing microbial diversity in the human microbiota. In 2005, Eckburg et al. analysed samples not only from faeces, but also from multiple colonic mucosal sites. They sequenced >13,000 16S rRNA genes, which constituted a substantial increase in scope over previous work, and discovered significant inter-subject variability and greater differences between stool and mucosal community composition than previously described.

Sequencing of marker genes (such as 16S rRNA) is traditionally associated with Sanger sequencing, which requires a labour-intensive

cloning step and is prohibitively expensive for large-scale microbiome studies. The advent of next generation sequencing (NGS) offered a cost-effective method that eliminated the cloning step by amplifying 16S rRNA genes using primers containing sequencing adapters and barcodes. The massive parallel sequencing throughput offered by NGS has significantly increased the sequence depth of 16S rRNA genes, allowing for taxonomic and phylogenetic analyses of complex microbial communities. Yet, 16S rRNA sequencing cannot always resolve closely related species and may miss the intra-species diversity. To better capture the full picture, shotgun sequencing was developed for direct sequencing of DNA. The advantage is its capability to recover the underrepresented microorganisms that were often masked by high-abundance species. Shotgun sequencing can use short-read (such as Illumina) or long-read (for example, Oxford Nanopore MinION and Pacific Bioscience Sequel) platforms. The short-read approach typically demands extensive computational support for assembling short reads into meaningful sequences. However, an

accurate and complete assembly of genome sequences is still hampered by the difficulty in resolving long repetitive regions. Thus, reference-based sequencing methods emerged.

Concerted sequencing efforts have been made to construct microbial reference sequences. In 2007, the National Institutes of Health launched the Human Microbiome Project and, five years later, published the first reference data for microorganisms collected from 242 healthy United States volunteers, covering a number of anatomical sites such as mouth, nose, skin, lower intestine and vagina (MILESTONE 17). The valuable genome references that were generated by this consortium allow reliable identification of individual microbial species, but fall short when new genomes are present in the community.

Long-read sequencing offers an alternative solution for mapping challenging repetitive regions. For example, single molecule, real-time (SMRT) DNA sequencing, in combination with a short-read shotgun DNA library, allows for de novo microbial genome assemblies, although it suffers from high error rates.

Access to genome sequencing has transformed human microbiome research from focusing on identity characterizations, to metagenomics approaches that not only reveal microbial species but also how microbial metabolic activities correlate with human health and disease.

In addition to metagenomics, metatranscriptomic analysis enabled by RNA sequencing offers a way to detect active members present in a microbial community (MILESTONE 14).

The construction of metagenome-assembled genomes, an approach also pioneered in environmental microbiology and recently applied to human-associated communities, will provide a new, unprecedented opportunity for deep characterization of the functional potential of the human microbial ecosystem.

Lei Tang, *Nature Methods*

ORIGINAL ARTICLES Wilson, K. H. & Blitchington, R. B. Human colonic biota studied by ribosomal DNA sequence analysis. *Appl. Environ. Microbiol.* **62**, 2273–2278 (1996) | Suau, A. et al. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Appl. Environ. Microbiol.* **65**, 4799–4807 (1999) | Kroes, I., Lepp, P. W. & Relman, D. A. Bacterial diversity within the human subgingival crevice. *Proc. Natl Acad. Sci. USA* **96**, 14547–14552 (1999) | Eckburg, P. B. et al. Diversity of the human intestinal microbial flora. *Science* **308**, 1635–1638 (2005) | Gill, S. R. et al. Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355–1359 (2006).

FURTHER READING Relman, D. A., Schmidt, T. M., MacDermott, R. P. & Falkow, S. Identification of the uncultured bacillus of Whipple's disease. *N. Engl. J. Med.* **327**, 293–301 (1992) | Hold, G. L., Pryde, S. E., Russell, V. J., Furrie, E. & Flint, H. J. Assessment of microbial diversity in human colonic samples

by 16S rDNA sequence analysis. *FEMS Microbiol. Ecol.* **39**, 33–39 (2002) | Pei, Z. et al. Bacterial biota in the human distal esophagus. *Proc. Natl Acad. Sci. USA* **101**, 4250–4255 (2004) | Bik, E. M. et al. Molecular analysis of the bacterial microbiota in the human stomach. *Proc. Natl Acad. Sci. USA* **103**, 732–737 (2006) | Fredricks, D. N. Molecular identification of bacteria associated with bacterial vaginosis. *N. Engl. J. Med.* **353**, 1899–1911 (2005) | Hyman, R. W. et al. Microbes on the human vaginal epithelium. *Proc. Natl Acad. Sci. USA* **102**, 7952–7957 (2005) | Gao, Z., Pei, Z., Tseng, C.-H. & Blaser, M. J. Molecular analysis of human forearm superficial skin bacterial biota. *Proc. Natl Acad. Sci. USA* **104**, 2927–2932 (2007) | Grice, E. A. et al. Topographical and temporal diversity of the human skin microbiome. *Science* **324**, 1190–1192 (2009) | Human Microbiome Jumpstart Reference Strains Consortium et al. A catalog of reference genomes from the human microbiome. *Science* **328**, 994–999 (2010).