GENOME WATCH

A glut from the gut: metagenomics takes a giant step forward

Alan Walker

Interest in the bacteria that inhabit the human gut has increased greatly in recent years, as technological advancements have allowed new discoveries. This month's Genome Watch reviews a recent study of the microorganisms in the human gut that sets a new marker for the field of metagenomics.

The human colon is home to a diverse collection of microorganisms, collectively termed the gut microbiota, which have profound effects on human health. Research into these microorganisms has been hampered by the complexity of the microbiota and by the strictly anaerobic nature of many of the constituent species, which renders them difficult to culture and characterise.

Metagenomics — the sequencing of DNA from an environmental sample — offers a powerful means to better understand the complex relationships between humans and their bacterial inhabitants.

This burgeoning technology was used in 3 pioneering studies to generate ~ 3 gigabases (Gb) of sequence from 33 Japanese and American individuals^{1–3}.

Now, Qin *et al.*⁴ have used Illumina sequencing technology to produce a metagenomic data set containing almost 200-fold more gut microbial sequence than was generated in the earlier metagenomic studies, a feat that would have been unimaginable just a few years ago. As part of the

MetaHIT project⁵, the researchers generated 576.7 Gb of sequence from

faecal samples from 124 European individuals, including healthy, overweight and obese adults as well as patients with inflammatory bowel disease. These sequence data were assembled into contigs, which were then used to build a total microbial gene catalogue containing ~ 3.3 million non-redundant predicted coding sequences (CDSs), which is ~ 150-fold higher than the number of genes encoded by the human genome. This catalogue is estimated to include a high proportion of all the microbial genes associated with the human gut, as shown by the fact that over 70% of reads from the three previous metagenomic studies also map to the contigs.

The depth of sequence coverage, even at these unparalleled levels, was not sufficient to detect all of the genes in each individual, but the authors detected ~ 500,000 unique CDSs per sample. Individuals must therefore share many bacterial genes from the total pool of 3.3 million CDSs, and the authors calculated that almost 40% of the genes from each individual were present in at least half of the other individuals in the same cohort.

Over 99% of the CDSs were of bacterial origin, and compositional analysis revealed 57 species that were common to more than 90% of the samples. However, in agreement with previous studies, there was a high variability (a difference of up to ~ 2,200-fold) in the abundance of these species between individuals.

Functional analysis of the gene catalogue identified a core set of genes that seem to be essential for bacterial survival in the gut. These include typical housekeeping genes and others that may encode products involved in adhesion to host proteins or in harvesting the sugars that are carried on blood or epithelial cells. Some functions were present in all 124 individuals included in the study cohort but detected in only a small proportion of previously-sequenced

genomes of gut bacteria. Therefore, the genes encoding these functions may be essential for the functioning of the gut ecosystem as a whole but may be encoded across different bacterial species. Although these CDSs included expected factors such as proteins involved in polysaccharide degradation and in the synthesis of short-chain fatty acids, amino acids and vitamins, a large proportion were poorly characterized, highlighting our inadequate understanding of gut microorganisms.

This study shows that Illumina technology, if used to a great enough depth, can generate meaningful sequence information from complex microbial communities, and it has created a fantastic resource for future research. However, the cost and effort involved in generating these data sets are considerable, and many of the initial conclusions are confirmatory rather than revolutionary. The great challenge now is to use this and subsequent resources to better characterize the gut microbiota and its impact on human health.

Alan Walker is at the Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK.
e-mail: microbes@sanger.ac.uk
doi:10.1038/nrmicro2363

1. Gill, S. R. *et al.* Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355–1359 (2006).
2. Kurokawa, K. *et al.* Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res.* **14**, 169–181 (2007).
3. Turnbaugh, P. J. *et al.* A core microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
4. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
5. Mullard, A. Microbiology: the inside story. *Nature* **453**, 578–580 (2008).

Competing interests statement

The author declares no competing financial interests.

FURTHER INFORMATION

MetaHIT: <http://www.metahit.eu>

ALL LINKS ARE ACTIVE IN THE ONLINE PDF